Auto-Carto Six

Proceedings of the Sixth International Symposium
on Automated Cartography

Les actes du sixième Symposium international
sur la cartographie automatisée

Volume II

AUTOMATED CARTOGRAPHY : INTERNATIONAL PERSPECTIVES ON ACHIEVEMENTS AND CHALLENGES

LA CARTOGRAPHIE AUTOMATISÉE : PERSPECTIVES INTERNATIONALES SUR LES RÉALISATIONS ET LES DÉFIS

Volume II

Edited by
BARRY S. WELLAR

Sous la direction de
BARRY S. WELLAR

October 16-21
National Capital Region
of Canada

Du 16 au 21 octobre
La région de la capitale nationale
du Canada

Published by:
The Steering Committee
for the Sixth International
Symposium on Automated Cartography

Publié par:
La comité organisateur
pour le sixième symposium
sur la cartographie automatisée

ISBN: 0-9691560-1-4 (Vol II)
ISBN: 0-9691560-2-2 (Vol I & Vol II)

Edited by/Sous la direction de

BARRY S. WELLAR

Published by:

The Steering Committee
Sixth International
Symposium on
Automated Cartography

Publié par:

La comite organisateur
du sixième symposium
international sur la
cartographie automatisée

CHAIRMAN'S FOREWORD

The Proceedings of the Sixth International Symposium on Automated Cartography are landmark in the documentation of the field of computer assisted cartography. None of the previous five sets of proceedings is comparable, as Auto-Carto Six has a number of unique characteristics which are reflected in this record of the symposium.

A special effort was made to emphasize international perspectives and the authors whose papers are presented here represent more nations than at any other Auto-Carto meeting. The proceedings also have a uniquely Canadian flavour, with a number of outstanding papers in both English and French. Canada has been in the forefront of the field of computer assisted cartography for the last two decades, and Auto-Carto Six provides Canadians with a unique opportunity to take stock of their achievements and to identify new challenges.

Although the editor of the Proceedings, Dr. Barry Wellar, has played a special role in their production, they would not have been possible without the assistance of all those who participated in Auto-Carto Six. The Proceedings of Auto-Carto Six are the written record of a remarkable collective endeavour.

D.R.F.T Taylor, Chairman
National Steering Committee
for Auto-Carto Six

September 27, 1983

PROGRAM DIRECTOR/EDITOR'S FOREWORD

The Proceedings are a single, but nevertheless very important indicator of what was said and done at any Conference. And, their value increases if a representative selection of papers can be assembled for the benefit of participants, as well as for interested parties not able to attend the meeting.

In the instance of Auto-Carto Six, it appears fair to say that the international Auto-Carto community outdid itself: one hundred and ten complete papers and fifty abstracts from speakers representing nineteen countries, including submissions in English and in French. I am grateful to the Technical Program Committee for assistance in securing papers, and to authors and their typists for submitting materials in the appropriate format. Their cooperation made distribution of the Proceedings in a timely manner possible.

It is appropriate in closing to acknowledge the efforts of several persons who not only made my job easier but improved the quality of Auto-Carto Six products and processes: David Douglas, Department of Geography, University of Ottawa, who was of invaluable assistance through his work as Special Sessions Coordinator, and as a resource person for the General Tracks as well; Johanne C. Forgues and Manon Forget, Department of Geography, University of Ottawa, and Doreen Ramplee-Smith, Carole Baker, Lloyd Mackey and Anne Buie of Carleton University, who prepared much of the material leading up to production of the Proceedings; and Bob Aangeenbrug, University of Kansas, Director of Auto-Carto 2, 3 and 4, Jack Foreman, (U.S.) National Oceanic and Atmospheric Administration, Director of Auto-Carto 5, and William French, American Society of Photogrammetry, who by their efforts set the production precedents for this publication.

September 27, 1983
Ottawa, Ontario, Canada

Barry S. Wellar
Department of Geography
University of Ottawa

PREFACE

Papers in the Proceedings are grouped to reflect the Symposium themes and sub-themes as follows:

A. Keynote and Plenary Addresses (Volume I)

B. General Track Sessions (Volume I)
   a) Integrated Systems
   b) Practical Applications
   c) Problem Analysis/Decision Support Systems
   d) Research and Development
   e) Education and Training

C. Special Track Sessions (Volume II)
   1. Limited Investment Mapping Systems
   2. Institutional Mapping Efforts: Status and Prospective
   3. Data Structures
   4. Mathematical Cartography
   5. Problem Analysis: Projects to Processes
   6. Thematic Mapping
   7. Remote Sensing: Systems and Activities
   8. Electronic Distribution/Exchange of Cartographic Data

The majority of articles presented in the Proceedings were assembled from camera-ready copy provided by authors. Abstracts are included in cases where complete papers were not submitted. In all events, editorial changes were kept to a minimum.

TABLE OF CONTENTS

VOLUME II

Chairman's Foreword iii

Program Director/Editor's Foreword iv

Preface v

Special Track Sessions 1
1. Limited Investment Mapping Systems
2. Institutional Mapping Efforts: Status and Prospective
3. Data Structures
4. Mathematical Cartography
5. Problem Analysis: Projects to Processes
6. Thematic Mapping
7. Remote Sensing: Systems and Activities
8. Electronic Distribution/Exchange of Cartographic Data

Appendices 629
Steering Committee
Technical Program Committee, Editorial Committee
Acknowledgements
Author Index

LIMITED INVESTMENT MAPPING SYSTEMS

The XYNIMAP Family of Systems for Geographic Information Processing and Thematic Map Production 2
David Douglas

Conversations with Odyssey 15
Martin Broekhuysen and Geoffrey Dutton

Un logiciel de cartographie assistée par microordinateur 25
Jean-Paul Donnay

LA CAD: A Computer-Assisted Design Package for Microcomputers 33
Robert Itami and Randy Gimblett

Small Area Mapping System (SAM) 44
Shiu-Yeu Li and John Morrison

Designing Interactive Cartographic Systems Using the Concepts of Real and Virtual Maps 53
Harold Moellering

Design of a TELIDON Based Image Analysis System (TELIAS) 65
Glen Shirtliffe

INSTITUTIONAL MAPPING EFFORTS - STATUS AND PROSPECTIVE

A Report on the Benchmark Testing of a Prototype Enhanced Stereoplotter Workstation 75
Rob John

Status of Auto Carto at the Defense Mapping Agency 85
F.C. Green

Une base de données relief a l'IGN France 90
D. Piquet-Pellorce et A. Bernard

Fabrication de bases de données à référence spatiale à partir de données numériques de la carte de base du Québec à l'échelle 1:1000 99
Gervais Pellerin

The Future in Terrain Elevation Data Processing at the Defence Mapping Agency 109
Robert Edelen

Cartographie assistée par ordinateur pour le recensement du Canada 117
Ross Bradley

Hydrographic Survey Requirements System (HYSUR) 127
Chung Hye Read

Computer-Assisted Cartography for Census Collection: Canadian Achievements and Challenges 135
Joel Yan and Ross Bradley

L'utilisation des systèmes graphiques interactifs en cartographie 147
Dominique Bruger

Automated Standard Nautical Chart Production 155
Larry Strewig, Joseph Ruys and Jan Schneier

A Stereo Electro-Optical Line Imager for Automated Mapping ... 165
J.R. Gibson, R.A. O'Neil, R.A. Neville, S.M. Till and W.D. McColl

Underlying Requirements and Techniques of a Geographic Data Base (Abstract) 177
Christine Kinnear and David Meixler

Kern MAPS 300 (Abstract) 178
Jonathan Hescock

Geographic Data Procesing in Forestry - Does it Pay? 179
J.A. Benson

Long-Term Plans for the Geographic and Cartographic Support to the U.S. Bureau of the Census (Abstract) 184
Frederick Broome, Timothy Trainor and Stephen Vogel

DATA STRUCTURES

Geodesic Modelling of Planetary Relief 186
Geoffrey Dutton

A Structured Expert System for Cartography Based on the Hypergraph-Based Data Structure (HBDS) 202
François Bouillé

Building a Hypergraph-Based Data Structure: The Examples of Census Geography and the Road System 211
Robert Rugg

Multiple Data Structures in a Regional Data Base 221
Paul Wilson

Adaptive Grids for Geometric Operations 230
Randolph Franklin

Terrain Approximation by Triangular Facets (Abstract) 240
Albin Tarvydas

Reduction of Digital Aerogeophysical Data to a Linear Model (Abstract) 241
Michael Holroyd

MATHEMATICAL CARTOGRAPHY

Scale Preserving Smoothing of Islands and Lakes 243
John Oomen and R.L. Kashyap

An Adaptive Method for Numerically Modelling Large Numbers of Irregularly Spaced Data 252
Richard Schiro and Glen Williams

About Cartographic Contouring with Computers 262
Pinhas Yoeli

A Mathematical Evaluation of Simplification Algorithms 267
Robert McMaster

Visual Versus Computerized Seriation: The Implications for Automated Map Generalization 277
Jean-Claude Muller and John Honsaker

Automated Detection of Drainage Networks from Digital Elevation Models 288
David Mark

Shape Representation by Rectangles Preserving Their Fractality 299
Pietro Dell'Orco and Marco Ghiron

Fractal Enhancement for Thematic Display of Topogically Stored Data 309
Marc Armstrong and Lewis Hopkins

Measuring the Fractal Dimensions of Surfaces 319
Mark Shelberg, Nina Lam and Harold Moellering

An Algorithm for Variable-Width Feature Representation (Abstract) 329
Barbara Bonnette

PROBLEM ANALYSIS: PRODUCTS TO PROCESSES

Digitizing the United Kingdom River Network 331
Roger Moore

Carte des sols et carte des terres agricoles 339
Jean-Phillippe Grelot

Cartographic and Attribute Data Base Creation for Planning Analysis Through GBF/DIME and Census Data Processing 348
Apollo Teng

Computer Mapping for Biomass Inventories 355
W.C. Moore and W. Chow

Flood Estimation in Europe: A Case Study of Applied Digital Cartography 365
Steven Wiltshire

A Geometric Mine Modelling System 374
Olli Karonen, Marku Tamminen, Penti Kerola, Matti Mitjonen and Esko Orivuori

Computer-Aided Design in Relation to Earthwork Computations .. 384
Stephen Mruk

Automated Data Base Capture for CADD 398
Richard Stover

Extended Graphical Plotting with PLANICOMP (Abstract) 408
Dierk Hobbie

MOSS - A State of the Union Address (Abstract) 409
Carl Reed

The Geographically Encoded String Structure (GESS) As Applied to the 1980 Census Digital County Boundary File (Abstract) ... 410
Roy Borgstede

Application of a Commercial Mapping/Data Management System to Forest Land Management in Maine (Abstract) 411
Daniel Boss

EPPL: An Innovative Grid Geographic Information System Language (Abstract) 412
Joseph Zastrow

A Three-Dimensional Model for Evaluating Potential Hydro-Geothermal Resource Utilization (Abstract) 413
Thomas Hodler

Some Problems in the Mapping of Geological Surfaces (Abstract) 414
Christopher Gold

THEMATIC MAPPING

Orthogonal Three-Dimensional Views for Thematic Mapping 416
Kurt E. Brassel and Zissis Kiriakakis

Automatic Colouring of Maps According to the Elevation 426
B. Falcidieno, C. Gambaro and P. Sinigaglia

A Procedure for Shading Class Interval Regions 435
Robert Cromley

Accuracy of Viewing Three-Dimensional Bars in Perspective: The Case of Computer-Generated Pillars as a Thematic Map Symbolism Type (Abstract) 443
John Smyrnew

A Program for Automatic Name Placement 444
J. Ahn and H. Freeman

The Making of the Far Eastern Economic Review Economic and Social Atlas of the Pacific Basin 454
D.K. Forbes, C.C. Kissling, K. Mitchell, M.J. Taylor, N.J. Thrift and S. Wild

Towards an Electronic Atlas 464
Ewa Siekierska

Application of a Model of Dynamic Cartography to the Study of the Evolution of Population Density in Spain from 1900 to 1981 475
J.A. Cebriân de Miguel

Le traçage automatisé en mode négatif: la couche à tracer et la couche pelliculable 484
Denis Genest

Computer to Map: An Exercise in Communication (Abstract) 491
Rosemary Ommer and Cliff Wood

REMOTE SENSING: SYSTEMS AND ACTIVITIES

The Analysis of Landsat Imagery Using an Expert System: Forestry Applications 493
M. Goldberg, M. Alvo and G. Karam

LANDSAT Digital Data for Updating Glaciological Information on Topographic and Glacier Inventory Maps 504
Philip Howarth and Simon Ommanney

Enhancements and Classifications of LANDSAT Data for the Ecological Resource Survey and Mapping of the Aishihik Lake Environmentally Significant Area (ESA), Yukon Territory, Canada 514
Dave Puittinen and A.B. Kesik

La cartographie thématique en télédétection 524
J.R. Bélanger

A High-Accuracy Airborne Digital Line Imager 536
R.A. Neville, S.M. Till, W.D. McColl and J.R. Gibson

Analytical Plotter for Facility Management Systems Data Acquisition 541
R.R. "Sky" Chamard

Underwater Mapping Techniques Using Remote Sensing for Salvaging Sunken Vessels 549
A.A. Navarro

Computer-Assisted Photo-Interpretation (Abstract) 559
Carl Reed

Automatic Cartography of Agriculture Zone by Means of Multi-Temporal Segmentation of Remote Sensing Images (Abstract) 560
Eric Cals, Robert Jeansoulin and J. Claude Darcos

Limitations of LANDSAT Imagery for Thematic Urban Mapping (Abstract) 561
Mohammad Id Ozone

ELECTRONIC EXCHANGE/DISTRIBUTION OF CARTOGRAPHIC DATA

An Intermediate-Scale Digital Cartographic Data Base for National Needs 563
Stephen Guptill

Computer-Assisted Mapping for the Census of Canada 570
D. Ross Bradley

The Availability and Use of Digital Topographic Data 580
M. Rodrigue and L. Thompson

Two Way Data Transfer Between Aries Image Analysis System and ARC-INFO Geographic Information System 588
Robert Maher, David Colville and Douglas Rigby

On the Transfer of Remote Sensing Classifications Into Polygon Geocoded Data Bases in Canada 598
D.G. Goodenough, G.W. Plunkett and J.J. Palimaka

A Demonstration Transfer of Remotely Sensed Data Utilizing the Standard Format for the Transfer of Geocoded Polygon Data 607
Kenneth Korporal

Cartographic Feature Coding 616
Bruce Wood and David Douglas

The Archiving of Computer Cartography (Abstract) 627
Louis Cardinal and Betty Kidd

LATE ARRIVAL(S)

IGDMS: An Integrated Geographic Data Management System (Abstract) 628
Timothy Nyerges

LIMITED INVESTMENT MAPPING SYSTEMS

The XYNIMAP Family of Systems for Geographic Information Processing and Thematic Map Production 2
David Douglas

Conversations with Odyssey 15
Martin Broekhuysen and Geoffrey Dutton

Un logiciel de cartographie assistée par microordinateur 25
Jean-Paul Donnay

LA CAD: A Computer-Assisted Design Package for Microcomputers 33
Robert Itami and Randy Gimblett

Small Area Mapping System (SAM) 44
Shiu-Yeu Li and John Morrison

Designing Interactive Cartographic Systems Using the Concepts of Real and Virtual Maps 53
Harold Moellering

Design of a TELIDON Based Image Analysis System (TELIAS) 65
Glen Shirtliffe

THE XYNIMAP FAMILY OF PROGRAMS
for
GEOGRAPHIC INFORMATION PROCESSING
and
THEMATIC MAP PRODUCTION

David H. Douglas
University of Ottawa

ABSTRACT

The XYNIMAP family of programs offers a wide range of cartographic and geographic information processing capability including XYNIMAP itself, an interactive digitizing and graphic editing system, BNDRYNET and CONSURF, programs for reorganizing data topology, and XYNIDRAW, XYNISYM, and VUBLOK, which are thematic mapping programs. Collectively the system illustrates what can be done utilizing timesharing and communication speeds which are considered as minimal for interactive text editing and other asynchronous remote terminal operations. The system was designed for all sorts of pilot and demonstration projects, to be a robust and flexible tool for undergraduate teaching, and as a research tool for geographic and cartographic projects that might be undertaken by faculty members who are not specialists in this field.

INTRODUCTION

It was in the universities in the late sixties, and the early seventies, that very much of the seminal development took place in the fields of computer assisted cartography and geographic information processing, especially in terms of concepts and algorithms. Yet it is ironic that most of the presently manufactured and marketed turnkey systems, (both large and small), which are part of the commercial expression of those early research efforts, are beyond the financial capability of the departments and schools that were so early involved. Geographic information processing is a tool which promises such an impact in the disciplines using maps it is imperative that students be introduced to it early in their post secondary careers if they are to keep up with the developments that will take place in their fields in the next few decades. It is a well known fact that evolution and development in most disciplines is profoundly affected by the available tools.

XYNIMAP

In response to a need for a simulation of a non-trivial interactive graphics work station environment for teaching, the XYNIMAP line digitizing system was developed exploiting the selective erasability, colour and flashing capabilities of a modern colour raster CRT display, and the available computing resources supplied by the University in the form of timesharing with a powerful central mainframe computer. Data output was to be easily massaged to a form required for commonly available geoprocessing and graphics systems, such as GIMMS (Waugh), the venerable SYMAP program which is still used because of the vast array of concepts it demonstrates (Doudenik), and innumerable programs obtained less formally than by commercial purchase. Modelled after the I.B.M. Scientific Subroutine Package which incorporates the documentation as detailed comments within the code itself, Douglas (1974) distributed a small geoprocessing subroutine library through the Harvard Laboratory for Computer Graphics and Spatial Analysis. The XYNIMAP family of programs, or system, represents a further progression, and perhaps a culmination of this latter work.

With the XYNIMAP program lines may be recorded as "reduced" streams of coordinates (reduced by the Douglas line reduction algorithm discussed in Douglas and Peucker, 1973), with identifying header, in files which may be stored, copied, archived, concatonated, or otherwise managed by any file handling system. Hardware components (Figure 1) include a digitizer or

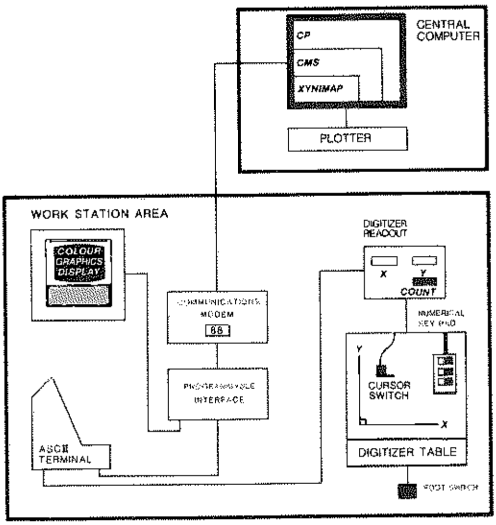


Figure 1. XYNIMAP Hardware Confuguration.

graphic tablet with point and time-stream modes of recording, a colour raster graphic CRT display terminal, a programmable interface, a modem and communication lines to a mainframe computer. XYNIMAP is a FORTRAN program which runs interactively with a graphics subroutine package for colour CRT terminals. The LOGITEK package (Karam and Raymond), with which XYNIMAP is currently implemented, is one of any number that would suffice, including the TEKTRONICS PLOT10 Library, and the Presentation Level Protocol (PLP) of TELIDON (O'Brien et al). XYNIMAP utilizes only the basic graphic functions from the display package, that is: change colour, fill a rectangle or polygon, draw a line, receive coordinates from a screen cursor, and write text. All scaling, windowing, and clipping functions are performed within the XYNIMAP program itself.

XYNIMAP mimics the actions of a "pencil and eraser", allowing the operator to: 1. trace digitize lines (at 3 points per second, a limitation imposed by the communication speed of 2400 baud), viewing them on the CRT screen almost immediately; 2. graphically "find" a line from a number of lines displayed with either the CRT or the digitizer cursor; 3. move, stretch, delete or break into shorter lengths, the line thus found; and 4. join pairs of lines. This activity is made clear to the operator by continuously updated menus displayed on a quarter panel of the graphic display, and by the use of selective erasability, signal colours and flashing with which the image is drawn. The graphic thus displayed matches what is being entered, changed or deleted from the data-set.

The coordinate values of points are transformed between the coordinate system of the digitizing table and that of the intended output coordinate system of the digital map, and again between the digital map and the graphics screen. The latter transformation is a simple scale and shift to a user defined window. The former is more complicated. By typing the coordinate values of "reference" points, and by digitizing these points from the graphic source document on the table, all other locations are transformed to the coordinate system represented by the typed values of the reference points. Three types of transform for this table-to-digital map conversion are implemented in XYNIMAP; namely: planimetric, projective, and conformal. Two reference points are necessary and sufficient to provide the parameters required for a planimetric transform. All digitized data are translated, scaled and rotated automatically, maintaining planimetric shape through the transform. Four reference points define a projective transform for which the code was taken from Baxter (pp. 133-134). The four points as typed need not be planimetric with those digitized, which implies a capability for shape change from the table to the output digital map. In fact a document may be stretched, squashed, inverted and distorted as desired, maintaining the property that straight lines remain straight. The conformal (or the Lauf Orphometric) transform maintains orthomorphism, or constant angles. The conformal transform parameters are calculated by the program with as few as three and as many as 100 reference points, although between four and eight are optimal. The multi-point conformal transform is especially useful for digitizing from the conformal projections, such as Mercator, Transverse Mercator, Lamberts, or Polar Stereographic projections. Since a conformal transform will convert straight lines to curves points are automatically inserted to ensure that curves result. If the user types the coordinates of reference points in latitude and longitude format, the output coordinate system is automatically set to the "Equatorial Mercator" projection. Data from other conformal projections are converted to that within tolerances required for a hand digitized product. Subsequent conversion to other projections directly, or indirectly through Plate Carree, is straight forward. Figure 2 illustrates the projection options.

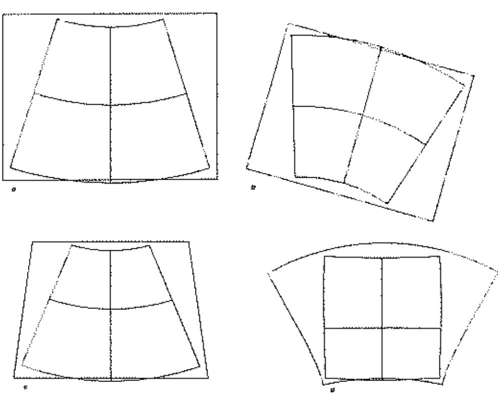


Figure 2. Transform Options: a) Input Lambert Conformal, b) Planimetric, c) Projective, d) Conformal.

Line data collected or generated by XYNIMAP are written into external files as binary numbers in variably spanned records with a header which allows the record to be reread by XYNIMAP, and the other programs in the family, without user involvement at the level of data formats, etc. This means that the actual numerical values, (that is the X,Y coordinates) are no longer a concern of the operator, who uses only the graphic form for his compilation or artistry.

COMMAND INTERPRETER

XYNIMAP, and the other programs in the family use a common command input interpreter which is similar to those found in GIMMS (Waugh) and ODYSSEY (Morehouse and Broekhuysen). The user can enter commands in a fairly free flow of words, numbers and carriage returns. For most command operations there is a liberal allowance of alias words. This is demonstrated by the sub-system in XYNIMAP which interprets locations entered as Graticule coordinates.

From time to time it is necessary for a user to enter locations in latitude and longitude coordinates. The problem is made more complex than normal interactive entry of numerical data because of the fact that entries will likely be made in an unknown selection of words and numbers representing axis, (ie. LATITUDE and LONGITUDE), direction, (ie. NORTH, SOUTH, EAST and WEST), and real values or integers representing DEGREES, MINUTES, and SECONDS, and it may be made in a variety of format conventions, (ie. longitude entered first). A specific module was encoded to interactively interpret longitude and latitude allowing for the above complications. The module calls the command analyser which receives the user´s list of words and numbers and then processes the axis, direction, degrees, minutes and seconds into radians. By means of displayed tables the process offers the opportunity to interactively correct and, if necessary, to re-enter the whole selection of points asked for by the XYNIMAP. The Latitude/Longitude reader routine utilizes the command parser and analyser, and its versatility is illustrated by the following set of valid responses for two required points.

```
LAT NOR 31 32 33 LON EAS 3 42 1 LAT NOR 34 35 36 LON EAS 5 6 10
NOR 31 32 33 EAS 43 42 41 NOR 34 35 36 EAS 46 45 44
31 32 33 43 42 41 34 35 36 46 45 44 *
LAT NOR 31 32 32 34 35 36 LON EAS 43 42 41 46 45 44
NOR 31 32 33 EAS 43 42 41 NOR 34 EAS 45 *
LAT NOR 31 LON EAS 43 LAT NOR 34 LON EAS 46 *
LAT 31 LON 43 LAT 34 LON 46
SERIES 31 32 33 34 35 36 43 42 41 46 45 44 *
31 32 42 43 34 35 46 44
31 43 34 46 *
LON EAS 43 42 41 LON EAS 46 LAT NOR 31 32 LAT NOR 31
NOR .031E+02, EAS 33.333 -13.6 -34.
```

The number of entries made per coordinate, per line of typing, etc., are compensated, and non-standard words and forms are analysed. All of the above twelve entries represent valid user responses to a program requirement to enter two points. Any of them may have been interrupted by one or more carriage returns and typed on a number of lines. Decimal, and exponential numerical values, including negative values, may be entered where integers are shown above. Complete words, or truncation to three letters, may be used for: LATITUDE, LONGITUDE, EAST, WEST, NORTH, SOUTH, and SERIES. Entries may be made without seconds, or without minutes and seconds, and since decimal numbers may be utilized, decimal degrees may be entered employing this feature.

The module has important recovery procedures for typographical errors. It displays a request for the number of points to the user and an example of a correct response. On receiving a blank line/carriage return, the routine first queries the user on words which it cannot interpret, such as the misspelled LITITUDE. On completion of this basic syntax recovery procedure a list of the interpreted and converted coordinate values is displayed. It then queries for confirmation providing the opportunity to retype any point, or it invites the user to re-enter the whole list if he chooses to do so.

One objective of the module's design was to make it unnecessary to trouble the user with detailed instructions illustrating the variety of ways to enter points. It is sufficient to instruct the user to enter LATITUDE and LONGITUDE as demonstrated by instructions issued by the module. The experienced user will very likely discover the possibilities on his own without the need for explicit instructions which, as can be seen, must be fairly verbose. In fact the philosophy behind the XYNIMAP system is that a minimum of rote learning from manuals should be necessary. Instead, input modules for all command sequences are designed to display acceptable actions as options, and robustly react sensibly to most other actions.

BNDRYNET

A number of informal experiments carried out suggested very strongly that the easiest way to hand digitize the boundary lines of a spatial partitioning is to follow the lines with the cursor without reference to the topology of the boundary network (see Douglas 1973). This applies to man-made networks, such as state and tract boundaries, as well as to representations of natural networks such as a soil-type partitioning. This is because at each node there is often one or more lines that pass right through without a change in direction at the node. For example it is possible to trace a single mid-continent line in the United States, from Canada to Mexico, which forms a part of the boundaries of eleven states, and passes through eleven network nodes, without a change in direction caused by the nodes themselves. It is easier to trace digitize this whole line in one pass, than to break it up in any other way. There are many other lines of this type forming this and any other network. Many of the lines seem to pass right through the nodes, almost as if oblivious of the topology, while the nodes are formed by other lines intersecting with them.

Because it is relatively easy to digitize simple lines a program was written to automatically recognize and record the topology of the polygon network implicit in them. BNDRYNET is, in other words, a cartographic spaghetti-to-polygon converter. Written in 1972, and possessing rather severe limitations in the number of segments, lines and polygons it can handle, it has nevertheless been incorporated as an output massager for XYNIMAP, but because of its limitations it is useful only to produce polygon perimeters for simple networks such as generalized representations of the U.S. states, the countries of Africa, or the census tracts of a small city. The program produces a line record for each polygon identified by a point digitized from the graphic within each desired polygon. Many programs and processes which employ this organization are available. Choropleth mapping programs for colour CRT plotters, for instance, are trivial to write if the polygons are recorded by their perimeters. Each boundary line which is output by BNDRYNET is, of course, made up of the same points as the boundary of the neighbouring polygon and there are no slivers or overlaps. A new algorithm now being encoded will provide a basis to eliminate these restrictions and output segment lists in a manner suitable for segment reading systems such as GIMMS. An efficient boundary network processor which produces a segment file without serious size limitations will form the foundation of a polygon network overlay system.

CONSURF

As discussed by Clarke, Gruen and Loon, contours are increasingly being used as sources of data for grid digital elevation models (DEM's). The contour map is practically the universal method for recording surface morphology in detail on paper maps, and they have stood the test of time for commensurabiity, utility, interpretability, and economy (of symbolism required for the information conveyed). A surface may be stored in digital form by line traces of the contours, but there are certain shortcomings related to what can be done with it in that form. On the other hand, a grid digital elevation model lends itself to many processes, operations, and to producing displays of all kinds. The reason for this obviously relates to its symmetry, but more especially to the implicit topological structure of the neighbourhood relations of each point. Nothing has to be searched to find which grid intersection point, or grid cell, or grid link is next to the current one, and data locations in storage are instantly calculable from the coordinate location in space. A contour to grid digital elevation model system seems an essential part of any integrated family of geoprocessing systems.

Conversion of structure from contours to a grid DEM consists of considerably more than reordering data elements on the basis of topology. Unlike conversion from line data (cartographic spaghetti) to a polygon structure, conversion from contours to grid involves a complete recalculation of all data element values. Hardly a single exact value will remain, except for the very occasional one where a point defining a contour falls exactly on a grid intersection. This is very rare. In general, therefore, all grid intersection values will be interpolated. The problem has two major components, a geometric one: that is the calculation of the value of a point from the values and locations of the contours near it, and an algorithmic one: the problem of finding those

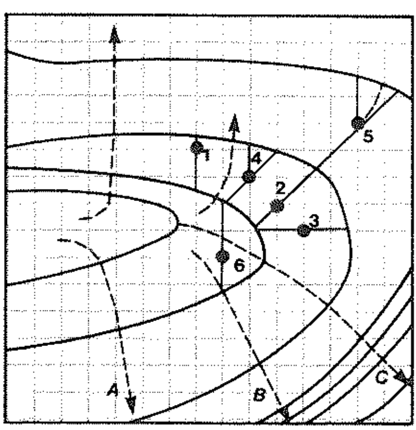


Figure 3. The Geometric Interpolation Component.

contours that apply, and the location along each contour to use for the interpolation, and further, to incorporate these in a tractable algorithm, given that for a problem of reasonably standard magnitude, 22,500 grid point values (ie. 150 by 150) might be calculated from a contour file comprised of as many, or more, than 150,000 points.

Figure 3 illustrates the geometric interpolation problem. In theory the value of a point must be interpolated along a slope line. Slope lines cross contours at right angles, and they converge to asymptotes along ridges and channels terminating at peaks or pits (see Warntz and Woldenburg, p. 57). A linear interpolation would probably suffice along the slope lines at A and B in Figure 3, however a four point polynomial interpolation, an overlapping pair of three point interpolations with a spline weighting function applied to the mid section, as suggested for line smoothing by Junkins and Jancaitis, or a third degree Hermite polynomial, suggested by Yoeli and discussed later, all merit serious consideration. These more sophisticated interpolations would likely reflect a more probable behavior of the surface in zones of transition such as evident along slope line C. Definition or calculation of a slope line from contours is very difficult however, and the gain by doing so will be marginal. For instance the value calculated at point 1 in Figure 3, by linear interpolation along the vertical grid line is probably close enough to an esoteric spline interpolation along a curving slope line that the difference can be ignored. This applies to points two and three as well, where the diagonal and horizontal grid directions are used to approximate the slope line. Points 4 and 5 indicate that a two direction approximation may even be closer to the curving slope line, and computations based on that may warrant consideration.

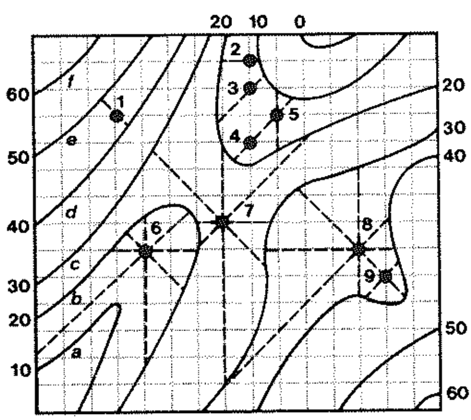


Figure 4. The Algorithmic Interpolation Problem.

Figure 4 illustrates some of the problem areas. At point 1 where there is little problem, it is evident that the value computed using the vertical, or the horizontal grid lines would not be very much different from that calculated along the Northwest-Southeast diagonal. The method suggested by Yoeli (1967) interpolates along a stratified sample of vertical grid lines using a three point Lagrange polynomial equation to find the slope at the centre of three consecutive contour intersections. A third degree Hermite polynomial is applied with the slopes at the contour intersections to interpolate to the grid points. A similar procedure is invoked to fill in the remaining points along the horizontal grid lines between the sample of vertical grid lines. The method lends itself to algorithmic expression, and the surface produced is generally smooth and of acceptable appearance. However there are serious directional biases that stem from the large angles between the approximated slope lines used, that is the orthogonals, and the actual slope lines. This shortcoming is criticized by suggestion in the text of Clarke, et al. though not specifically stated. At point 1 in Figure 4 it can be seen that the Northwest-Southeast diagonal can easily be isolated by a variety of tests, including a simple length test of the grid orthogonal or diagonal joining contours (d) and (e), length of the shortest "half" ray, that is the Southeast ray, or along the ray yielding the maximum slope. Clarke, et al, suggest that selection from four grid lines, that is the orthogonals and the diagonals, is superior. A line more closely paralleling a slope line can obviously be better selected from four choices than from two. Values computed for points 2, 3, 4, and 5 indicate changes of choice of grid directions along which to interpolate, and in fact, discontinuities occur in any surface calculated through such transitions. Point 6 illustrates a problem in extended contour loops such as along ridges and channels. Point 6 would be calculated as the value of contour (b), because none of the orthogonal, or diagonal rays necessarily intersect another contour. Points 7 and 8 demonstrate other difficult cases, the former the problem of generating a decent surface behavior in passes, and the latter where contours do peculiar things, such as looping in and out in opposite directions, when the expectation is that they remain more or less parallel. The only simple case, therefore, occurs when contours are close to parallel, and are relatively straight. At all other locations, which seem to be in a majority on most contour maps, surface irregularities and discontinuities result from all of the procedures so far discussed, and would occur for any other interpolation method, even if done along the exact slope line regardless of how it curves. A method for dealing with these discontinuities will be discussed below.

The algorithmic problem can be very succinctly stated. For a small grid measuring 150 by 150 grid cells, 22,500 values must be computed. For each one a search for the eight intersections of contour lines closest to the point along the four radiating grid lines must be conducted. This search must consider all the contours that pass over the 150 by 150 grid patch, and such a file can easily be composed of 150,000 points. Without a careful structuring of the process this could involve several hundred billion operations just for the search.

The objectives of CONSURF were defined as follows. Since the contour file is likely to be large the search through it for intersection points must be restricted to one pass. Further, as always, there is a requirement to use as little core as possible. Core allocated for CONSURF is restricted to three times the core required to store the grid itself. An algorithm based on a linear insertion list was devised. Arrays are created with each element to act as an independent entry root to the list, one for each grid line (Figure 5). Each segment of the contour file is examined to see

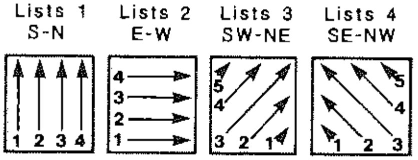


Figure 5. Superimposed Grid Line Lists.

what grid lines it crosses, if any, and for each one it does cross the Z value of the contour, is inserted into the list, along with, and sorted on, the Y value of the intersection with the grid line. In the case of horizontal grid lines, the X values are used. The root element for entry to the list is calculated from the segment coordinates. The multiple root system actually makes the one single list behave as though it were many individual lists, but because only one "NEXT" pointer is used for record entry, they are all, in effect, superimposed. This list, which can be quite large, may be written into direct access external storge, or, if available, may be kept in core.

When the search and the list are complete each grid line of intersection points may be reconstructed by retreiving from the list starting at the appropriate root entry. Grid point values are calculated along each grid line for each set of grid lines: all the horizontal lines, the vertical lines, and the two sets of diagonal lines. Two values are stored for each grid point, a Z value, and a D value. Each grid point is interpolated four times, and depending on the D value, the previous Z value and D value are retained, or replaced. There is a choice of what to use as the D value. It could be distance between successive intersection points in the XY plane, the slope, or the half ray length. A number of tests suggest that the half ray length was more useful, although the differences were marginal. To assist in making a better choice of grid ray in extended contour loops, and in passes, the D value computed from a horizontal interpolation segment is multiplied by four to make it less likely to be used over that computed from a sloping segment.

The surface that results from these algorithms represents the linear interpolated value along the grid line selected from the four on the basis of a contour intersection being closest to the grid point. This is modified, as discussed above, if the slope is zero along this segment.

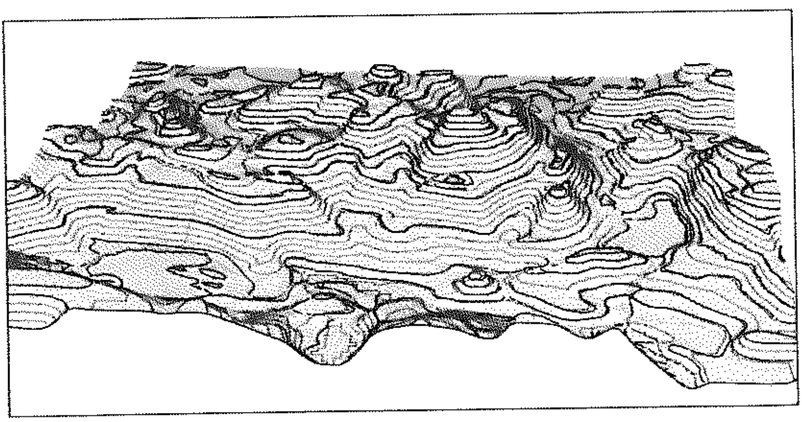


Figure 6. Contours interpolated from a Grid, which in turn was Interpolated from Digitized Contours.

The resulting DEM is predictably quite well behaved in areas well defined by parallel, relatively straight contours. The discontinuities mentioned above are frequent in other areas. These are dealt with by a selective local smoothing operator. The D grid, generated and stored along with the Z grid, records the distance from each grid point to the nearest contour.

As Tobler notes (p. 179), edge effects creep inward towards the centre of a geographic matrix with each consecutive local smoothing operation. In the same way, the smoothing effect diffuses outward from grid points protected from being smoothed no matter how this may be done. Since the D value records distance from the nearest contour, it provides a convenient indication of those points further than a grid unit from one. Only these points are smoothed, but the smoothing operator considers all neighbouring points, whether they are smoothed or not. In this way the fidelity of the representation in the vicinity of the contours is retained. Careful visual examination of perspective view plots of surfaces produced by these algorithms confirm that they produce an accurate reflection of the surface represented by the contours, and a very well behaved surface in regions between contours.

Figure 6 illustrates the contours from which a grid digital elevation model was interpolated, the grid lines, and contours computed from the grid.

GRAPHIC ROUTINES

A prerequisite of any geoprocessing system, especially for teaching, is a considerable graphic output capability. If data are well organized, and not too voluminous, thematic mapping programs, especially of the statistical map variety such as choropleth maps, proportional sized symbol maps etc. may be written with relative ease. Difficulties occur only when hidden line operations for line drawing plotters must be incorporated.

The XYNIMAP family at present offers three graphic options: XYNIDRAW, XYNISYM and VUBLOK.

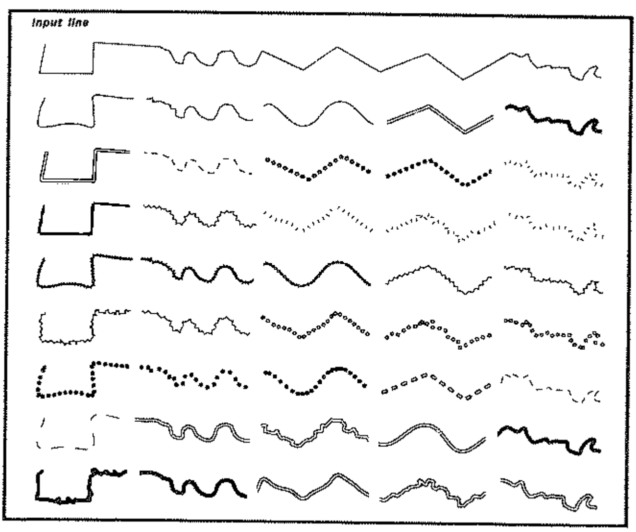


Figure 7. XYNIDRAW Line Symbolism Sample.

XYNIDRAW consists of a command input interpreter that allows identification of line features to which a line symbol type may be applied, along with scaling and window options. Line symbols may be created with permutations of line type commands selected from: WIGGLE, JAGGED, SMOOTH, DOUBLE, SOLID, BOXED, DASHED, DOTTED, SPOTTED, TRACKS, ZIGZAG, and DIAMOND. Each line type command has one or more numerical parameters to indicate such things as size. For example, SOLID may be followed by two real numbers giving the outer and inner widths of a solid line, which results in a hollow solid line. Figure 7 illustrates a sample of several hundred permutations that have been tested.

XYNISYM combines the capability to produce PILLAR maps (Douglas, 1979) and Proportional Circle Maps, (Rase, 1973).

VUBLOK is a perspective view display system which illustrates grid digital elevation models with a variety of symbolism types (Douglas, 1972). Surfaces may be illustrated in perspective view with plastic like relief shading if a raster display with a greyscale capability is available, profile lines, and horizontal plane contour lines. It is also possible to plot other input lines on the surface. These lines may be entered with fixed Z values associated with the lines themselves, such as shore lines and contour lines, or they may have their Z values computed from the model. The line symbols are plotted on visible zones of the surface only, which is of course dependent on the projection of the image and the relief of the surface itself.

SUMMARY

The XYNIMAP program, with its component software and hardware, acts as an automated pencil and eraser to record line data from graphic source documents into a previously defined coordinate system. A common configuration of the system (taking into account the C.P.U. core storage, and the direct access storage that can be allocated at a particular installation), will enable 1,000 lines of 1,000 points each to be recorded, edited, and stored in a single session. Lines are recorded without explicit structured relations to each other (cartographic spaghetti), and certain structuring procedures and data massage operations are available through the Boundary Network Program (BNDRYNET) and the contour to grid program (CONSURF). These programs are integrated on the basis of common data requirements and output, as are the graphic plotting and display programs with which a wide variety of thematic map types can be produced.

Perhaps more importantly, the development of this system has resulted in several algorithms including a new hidden line algorithm for digital elevation model surface displays, which is actually an optimal combination of three contrasting methodologies, an efficient polygon window clipper, the contour to grid procedure, and several others. With the XYNIMAP family some work has been carried out on the cartographic feature code problem, and on surface compression techniques.

ACKNOWLEDGEMENTS

The author wishes to acknowledge Thomas Wray of the Dept. of Energy Mines and Resources, Canada, for interpreting the Lauf Orthomorphic Transform. Funding for important enhancements to the VUBLOK system was generously granted by The Department of Energy Mines and Resources, Canada.

REFERENCES

Baxter, Richard S., 1976, Computer and Statistical Techniques for Planners, Methuen and Co. Ltd., London

Clarke, A.L., Gruen, A., and Loon, J.C, 1982, "The Application of Contour Data for Generating High Fidelity Grid Digital Elevation Models", Proceedings of Auto-Carto 5, ACSM/ASP Conference Crystal City VA, pp. 213-222

Doudenik, James A., 1975, "SYMAP User's Manual", Laboratory for Computer Graphics and Spatial Analysis, Harvard Graduate School of Design, Cambridge Mass.

Douglas, David H., 1972, "VIEWBLOK: A Computer Program for Constructing Perspective View Block Diagrams", La Revu de Geographie de Montreal, Vol. XXVI, No 1. pp 102-104

Douglas, David H., 1973, "BNDRYNET", in Peucker, T.K., Interactive Map in Urban Research: Final Report after Year One, University of British Columbia, Vancouver

Douglas, David H., and Peucker, Thomas K., 1973 , "Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or Its Caricature", Canadian Cartographer, Vol. 10, No. 2, pp.112-123

Douglas, David H., 1974, Collected Algorithms, Paper No. 20, distributed by the Laboratory for Computer Graphics and Spatial Analysis, Graduate School of Design, Harvard University

Douglas, David H., 1979, "The Pillar Mapping Program", in Harvard University Mapping Collection, vol. 2, Mapping Software and Cartographic Data Bases, Laboratory for Computer Graphics and Spatial Analysis, Harvard University, pp. 51-62

Junkins, John L., Jancaitis, James R., and Miller, Gary W., 1972, "Smooth Irregular Curves", Photogrammetric Engineering, Vol. 38, pp. 565-573

Karam, Gerald, and Raymond, Jacques, 1981, VM/LOGITEK User's Guide, 3rd. Edition, Dept. of Computer Science, University of Ottawa.

Lauf, G.B., 1970, "The Orthomorphic Transformation of Co-ordinate Systems", Proceedings of the First Canadian Symposium on Mining Surveying and Rock Deformation Measurements, Canadian Surveyor, Vol. 24, No. 1., pp. 238-245

Morehouse, Scott and Broekhuysen, Martin, 1982, ODYSSEY User's Manual, Laboratory for computer Graphics and Spatial Analysis, Cambridge MA.

O'Brien, C.D., Brown, H.G., Smirle, J.C., Lum, Y.F., and Kukulka, J.Z., 1982, TELIDON, Videotex Presentation Level Protocol: Augmented Picture Description Instructions, Communications Research Centre, Canada Dept. of Communications, Ottawa

Rase, Wolf D., 1980, Subroutines for Plotting Graduated Symbol Maps: Description and User's Manual, Version 1.2E, EDV Report 2, Bundesforschungsanstalt fur Landeskunde und Raumordnung.

Tobler, Waldo R., Ed., 1970, Selected Computer Programs, Dept. of Geography, University of Michigan, Ann Arbor

Warntz, William and Woldenburg, Michael, 1967, "Spatial Order: Concepts and Applications", Harvard Papers in Theoretical Geography, No. 4, Graduate School of Design, Cambridge

Waugh, Thomas C., 1975, GIMMS Reference Manual, Program Library Unit, University of Edinburgh, Inter-University Research Council Series

Yoeli,P., 1967, "The Mechanization of Analytical Hill Shading", Cartographic Journal, Vol. 4, No. 2, pp. 82-89

CONVERSATIONS WITH ODYSSEY*

Martin Broekhuysen and Geoffrey Dutton
Laboratory for Computer Graphics
and Spatial Analysis
Graduate School of Design
Harvard University
Cambridge, MA 02138 U.S.A.

ABSTRACT

ODYSSEY is a software system for the management, analysis and display of geographic information. It consists of nine Fortran programs, each devoted to one or more related tasks, including map digitizing, editing, transformation, verification, overlay and cartographic display. While the ensemble of tasks the system can perform is large, ODYSSEY can be understood and operated by relatively untrained individuals. This is no accident: man-years were spent in the development and refinement of command protocols and dialects which are simple to understand and consistent in the way they operate. This paper concentrates on describing the user command language: its logic, features and uses. Following a discussion of the philosophy of control embodied by ODYSSEY's language system, command language syntax is summarized, with examples (principally from the POLYPS mapping program) demonstrating both the consistency and flexibility built into the language system. Examples of user-program interaction are presented which illustrate the options available for controlling ODYSSEY programs and the forms of feedback presented to users, including informational, warning and error messages. Finally, command procedures are discussed, concluding with a description of ODYSSEY's interactive tutorial capability, a relatively simple extension of command file processing which provides users with on-line instruction upon demand, at the same time that they are in command of the program.

INTRODUCTION

The ODYSSEY system comprises a rich set of tools for the manipulation of two-dimensional cartographic data. Using the full set of tools, a user can digitize the coordinate and topological data defining the boundaries of a two-dimensional area coverage of a region; process the resulting file to make it geometrically accurate and topologically consistent; overlay it with one or more pre-existing similar cartographic files for the same region; perform a variety of common cartographic operations on the map file such as boundary generalization, coordinate scaling, translation or rotation, and polygon selection and aggregation; create

---

* The ODYSSEY language system was built by a number of researchers at the Laboratory for Computer Graphics and Spatial Analysis. The design and implementation of the system was done by Nicholas Chrisman, James Dougenik and Bruce Donald, with contributions from Denis White, Mark Kriger, Geoffrey Dutton and Thomas Jaskiewicz. The POLYPS language was designed and implemented by Scott Morehouse. The TUTOR language system was designed by Geoffrey Dutton, and implemented together with Scott Morehouse. Early research leading to the Laboratory's language system was supported in part by NSF grant MCS-74-14437-A01. An extended version of this paper was presented at Harvard Computer Graphics Conference, Cambridge Mass., 1983.

estimates of statistical attribute values for the polygons of a coverage from the values for polygons of a related coverage; and draw thematic maps of the resulting value distribution in black-and-white or color, and in either two-dimensional or three-dimensional perspective form, on a variety of display devices.

These capabilities are implemented in the nine programs of the system. Each program is invoked as an executable module called from the resident operating system; each program exits to the operating system under user control. The programs communicate via sequential files of cartographic data and statistical attributes in well-defined formats.

A single program comprises a functionally coherent set of capabilities within the entire system. Common applications of ODYSSEY call on several of the programs as a matter of course. A user may routinely use three or more of the programs in the course of half a day's work, returning to each of them more than once. For example, a standard demonstration of ODYSSEY capabilities consists of treating this problem: "Given an on-line map file of the boundaries of counties of the state of Wyoming; a corresponding attribute file of population statistics for these counties; and a paper map of the three-digit zip-code areas of Wyoming. Create an on-line map file of these zip-code areas by tracing them on a tablet; generate estimates of the population statistics for these areas; and draw a thematic map of population density by zip-code area." A solution of this problem invokes six of the nine ODYSSEY programs: HOMER and PENELOPE to create the map file of zip codes, WHIRLPOOL and CALYPSO to generate the estimated population statistics, CYCLOPS to create a polygon zip-code file, and POLYPS to draw the map.

As we mean to describe in this paper, the dialog with the user is consistent in form and meaning from one program to the next. The communication rules are uniform, and the programs' behavior is highly predictable. This is an important advantage to users of a large, multi-function application system.

It will be seen that the dialog is user-driven. There are no menus, no gratuitous prompts; and programs pose almost no questions to the user. The medium for interaction, and the means by which the user controls the program, is the ODYSSEY language system. It forms the user-program interface from the time the program is invoked, except during input from a digitizing device and during graphic output to a display device. Its functions can be described as follows:

(1) to prompt for commands;
(2) to parse user input;
(3) to assess the validity of each command;
(4) to report lexical and syntactic errors in the user input;
(5) to display the values of parameters;
(6) to initiate data input, processing and output on demand; and
(7) to process command files and conduct tutorials.

Digitizing and drawing are initiated by commands; and program-generated messages appear bracketed between items in the dialog with the language system. Evidently the language system holds the central place in ODYSSEY as perceived by the user. Our concern in this paper will not be with the exact structure or the meaning of ODYSSEY language commands, but with the user-program dialog as mediated by the language system.

INTERACTION WITH THE LANGUAGE SYSTEM

When an ODYSSEY program is invoked from a user's interactive terminal, it is ready to accept all command sentences. This initial state is signaled by the single-character prompt "?". This prompt is used by all programs when awaiting new commands, and cannot be disabled. POLYPS identifies itself, then awaits a command (here and in the following, user input will be shown indented in lower-case letters, and language-system responses in capitals):

```
        polyps

+++      WELCOME TO POLYPS
+++  ODYSSEY CHOROPLETH MAPPING

?
```

Commands are typed as strings of 80 characters or less, terminated and transmitted by a carriage return. One or more commands can be typed on a line. After the program has processed a line of complete, correct commands, the language system returns to initial state.

Commands consist of verbs, object names, modifying phrases, values (or expressions representing values), and punctuation. Each is terminated by a semi-colon, as these examples from POLYPS show:

```
        classify;
        include map overlay;
        zoom in: 2;
```

An input line is read, then analyzed character by character. In doing so, the language system analyzes value expressions and expands word abbreviations, checking each lexeme against the stored table of legal constructs. A legal command, terminated properly by ";", is parsed, acknowledged, and, if appropriate, executed immediately.

If an illegal lexeme or expression construction is encountered in the input line, the language system discards the remainder of the string, issues an error message to the terminal, and prompts the user with the partial legal command, if any:

```
?
        make key continuous;
WHAT? - UNRECOGNIZABLE SYNTAX
?MAKE KEY
```

In this interchange, MAKE KEY is a legal beginning of a command; but the word CONTINUOUS cannot follow MAKE KEY. At this point -- following the prompt -- the user can continue with the correct language:

```
        type continuous;
.-. MAKE KEY TYPE
.-. CONTINUOUS
?
```

ODYSSEY acknowledges input by displaying echoes (identified by the prefix ".-.") of the processed input, from the beginning of the command through the last complete, correct semantic unit. This language-echo function is optional; the experienced user can disable it to speed the interactive process. The new user may call on it to reinforce learning how words within commands can be abbreviated:

```
        ma ke ty cont;
.-. MAKE KEY TYPE
.-. CONTINUOUS
?
```

Any word may be abbreviated by its first few letters. The minimum abbreviation is the set of letters needed to distinguish the word from any other word which would be legal in the same context.

Typing "?" prompts ODYSSEY to display to the user a list of the legal next words (each prefixed by "==="), and then to prompt as above:

```
?                                   t ?
        m k ?                   === CONTINUOUS
=== MAXTICKS                    === NOMINAL
=== SIZE                        === ORDINAL
=== COLOR                       ?MAKE KEY TYPE
=== FONT                                c ?
=== TYPE                        === ;
=== NDECIMAL                    ?MAKE KEY TYPE CONTINUOUS
=== PENSIZE                             ;
=== POSITION                    .-. MAKE KEY TYPE
?MAKE KEY                       .-. CONTINUOUS
                                ?
```

In this manner new users can explore the command language tree to any level, identifying their choices and not committing themselves until they intend to.

An incomplete command line already accepted by the system can be cancelled by typing "!" as the first character of the next line:

```
        ma ke co;
WHAT? - UNRECOGNIZABLE SYNTAX
?MAKE KEY
        !
?
```

The partial command MAKE KEY is forgotten, restoring initial state.

The dialog involving command language and the special character "?" is nearly symmetric between program and user. As input, the language amounts to prompts to the program to do certain processing, and the "?" prompts the program to display a list of the possible next inputs. Output lines beginning with "?" are prompts to the user for further input. This behavior encourages a smooth and responsive dialog between ODYSSEY and its users. During the dialog, a user is always prompted with the current lexical state and optionally with the legal next inputs; and is empowered to commence, complete, correct, and cancel commands.

SEMANTIC ENFORCEMENT OF COMMANDS

Error messages, warning messages, and "+++" informational messages are generated not by the language system, but by the other processing modules attempting to carry out correct commands. Any of these may be triggered when a syntactically legal command asks for unusual or impossible results. The message is displayed following the language echo. The interaction following such a message depends on the sort of command. An error message is followed by a return to initial state. For example, this happens when an input data file can't be found:

```
        open base filename: 'us.pdg';
.-. OPEN BASE FILENAME: "US.PDG"
ERROR - CANNOT OPEN FILE "US.PDG"
        PLEASE CORRECT ERROR AND TRY AGAIN
?
```

Similarly when a parameter is assigned a semantically invalid value:

```
        make classification nlevels: -7;
.-. MAKE CLASSIFICATION NLEVELS:  -7
ERROR - NLEVELS MUST BE GREATER THAN 0 AND LESS THAN 21
        PLEASE CORRECT ERROR AND TRY AGAIN
?
```

The language system's behavior in the case of a warning is of two kinds. These correspond to the two general kinds of commands: commands to manipulate program parameters, and commands to process files. A command which manipulates a parameter in a dubious fashion may have unwanted results; this produces a warning message and a return to initial state:

```
        make viewport size: (13,10);
.-. MAKE VIEWPORT SIZE: (13, 10)
WARNING - VIEWPORT IS PARTIALLY OUTSIDE THE PAGE LIMITS
```

Here, a probable oversight has been noted by POLYPS, one that can be corrected by simply re-specifying the page size or the viewport size. Potentially more serious, a command to process a file could result in expensive, useless results, due to certain parameters having not been specified. In many instances, ODYSSEY can warn users about the situation. The user is given the option of aborting the processing command:

```
        run;
.-. RUN
WARNING - NO ERRORREPORT FILE
+++ TYPE C TO CONTINUE
    TYPE Q TO RETURN TO THE COMMAND LANGUAGE
```

In this dialog with the WHIRLPOOL program, the displayed warning reminds the user that potential errors from the run will not be documented because no report was requested. Every program uses the "C or Q" prompt to give users the chance to begin or abort processing when potentially adverse consequences have been identified.

COMMAND-FILE INPUT

Typing READ COMMAND FILENAME: "name"; directs the language system to immediately begin accepting command text from the designated file rather than from the user's terminal. If it can be found and opened, the file is read and executed, line by line. Upon reaching end-of-file, control is returned to the user, or to the prior command file. As many as ten command files can be stacked (left open simultaneously). While this may seem unessential, this capability to "layer" command files is regularly exploited in "tutorial" command mode, discussed below.

During command file processing, ODYSSEY programs act consistently, with adaptations in recognition that dialog is impossible with canned instructions. As before, output from the language system and the rest of the program is directed to the terminal, including language echoes, program warnings and error messages. Furthermore, should syntax or certain other error conditions occur, subsequent commands to DRAW, COPY or otherwise process files are acknowledged but ignored, as is standard practice in batch processing. In ODYSSEY sessions, command-file errors may be rectified (or ignored) by the user, who can elect to reissue the directive to process, which the program will then obey.

Command files can also include comments to document their purpose and explain choices of parameters, authorship, etc. Comments are any text following the character $ (dollar sign); all text following this flag is read and echoed (if echoing has not been disabled), but not otherwise interpreted by the language system. (One exception: character strings, such as Titles and Text annotation, can include dollar signs, which will be plotted as such.) The comment feature, more or less an afterthought in the design of the system, turned out to provide a basis for implementing on-line tutorials, described next.

ON-LINE TUTORIAL FACILITY

A potential deficiency in ODYSSEY's user interface is the lack of "help" messages. While the language system could feature a HELP command, the complexity and diversity of ODYSSEY would be hard to describe in simple messages. Since the meaning of commands such as DO, MAKE, CREATE and SHOW is fairly colloquial in ODYSSEY, the kinds of help users might want will normally be at higher conceptual levels, involving explanations not of lexemes but of techniques. Assistance of this nature has been made available as a prototype tutorial for the POLYPS mapping program. Rather than being built into POLYPS, it resides in a set of files accessed by the READ COMMAND FILE command and managed by a special interpreter, common to all ODYSSEY programs.

Tutorials are specially structured command files, containing both commands and comments, with comments of certain types acting as instructions to control the structure and progress of the tutorial. Normally, when executing a command file, control is transferred to the file (away from the user) until the file issues a QUIT command (ending the session) or reaches its end (returning control to the user). If that file is conducting a tutorial, however, it can be made to pause at any line to let the user issue commands, either to the tutorial (to continue, select, list, or skip) or to the program itself, as usual: users always have a choice between issuing commands and getting information and examples; they are never force-fed with "computer-aided instruction". More important, the ease of structuring tutorial files empowers users to produce their own for the benefit of others.

ODYSSEY's language system includes a facility to recognize tutorial command files and manage interaction between them and users, the TUTOR subprogram. It behaves appropriately when a tutorial file opens an ordinary command file, as well as when a lower-level file is itself another tutorial. When several levels of tutorials are active, the effect is like a "menu", from which users can select topics, browsing up and down the hierarchy. The POLYPS tutorial, supplied with ODYSSEY, actually consists of several dozen command files organized in a tree structure, plus a set of cartographic files used to formulate examples. Users need not be aware of what or where these files are, but do need read-access to all of them.

To activate this tutorial is the decision of the user, who is informed of that option when first invoking POLYPS. The following message, stored in a "profile" command file (which POLYPS opens and executes automatically upon startup) is issued:

```
      exec polyps

+++        WELCOME TO POLYPS
+++   ODYSSEY CHOROPLETH MAPPING

      DEFINE HELP:'PTUTOR.CMD';
$
$    Help is available on many topics in POLYPS.
$    To receive it, please type the following command:
$
$     READ COMMAND FILE:HELP;
$
$    You will be instructed then on how to ask for
$    more specific information.  Back to you.
$
?
```

If at any point during the execution of POLYPS the user asks for this help, the following occurs:

```
      read command file:help;          <User invokes tutorial>

+++ BEGINNING TUTORIAL

   You are now engaged in the POLYPS tutorial.  All program
   commands will still be executed,  but you may also consult
   any of the available topics for discussions and illustrations
   of how to control POLYPS and what you can do with it.
   There are 6 special commands for using the tutorial:

  .    -proceed to the next paragraph;
  .;   -skip to the next topic;
  .!   -get the next more general topic;
  .#   -list current set of topics;
  .:N  -select topic N;
  .?   -review these commands;

      As distributed by Harvard, this Tutorial for POLYPS has
   eleven major topics, all of which have many subtopics containing
   instruction, hints and demonstrations.  The set of topics
   available to you changes as you enter into greater detail.
   Use the ".#" command to list current ones.
```

```
    Here we go...

 1   Getting Started                          <First topic announced.>
 .?                                           <Control returned to user.>
```

At this point, the user is free to continue with the tutorial or issue commands to POLYPS. Typically, one is interested in learning what is available in the tutorial, and would request its table of contents:

```
 .#                                           <User asks for topic list.>

 1   Getting Started
 2   Using the POLYPS Command Language
 3   Data Files Used by POLYPS
 4   Understanding Graphic Display Devices
 5   Map Elements:  Their Properties and Uses
 6   Geo-graphic Formatting:  Manipulating Base Maps
 7   Classification:  Interpreting Thematic Variation
 8   GREYSCALE Maps:  Black & White and Graphic all over
 9   MYSHADE:  A Symbolism Kit
 10   NOMINAL Maps:  Communicating Categorical Data
 11   Using Color Symbolism and Annotation
```

The user can select any of these subjects (using the tutorial command .:) or simply wade in by typing a period, if just getting started.

Consistent with ODYSSEY's overall approach, the interaction strategy adopted for tutorials is to keep the user in control, with the program passive, but responsive. Suggestions are offered, but specific responses are neither demanded nor evaluated. There is, in fact, no mechanism for processing feedback, not even a simple YES/NO branching operator. Consequently, tutorials explain and illustrate, but do not attempt to converse. They provide consultation, "over-the-shoulder" so to speak, allowing users to peruse the tutorial or issue ODYSSEY commands, without interference and at their own pace.

Field Production of Tutorials

Its consistency, interactivity and power notwithstanding, the ODYSSEY language system is formally, immutably specified by its designers and cannot readily be modified in the field. In contrast, the tutorial facility is an open-ended tool, available for documenting the devices, databases, procedures and projects at any installation. By adding commentary and embedding certain control flags, one can expand any command file into a tutorial. These flags must be positioned at the beginning of records, which must be identified as a comment. That is, its first character must be a Dollar Sign ($). A command file is identified as a tutorial if and only if its first two characters are "$#". In addition to waking up the TUTOR subprogram, this record also contains the name of a file which lists the topics of the tutorial. This menu is printed when the user types ".#".

Another comment-command, "$/", must terminate every tutorial file. Unless the last record of a tutorial contains these two characters, the pre-tutorial environment may be incorrectly restored. Once bracketed by these sentinels, a command file is a tutorial, at least formally. Four other commands can then be interspersed at will to annotate and structure the file. These control statements commence with the characters "$$", "$?", "$." or "$;", and function as follows:

$$	Flags text to be printed
$;	Flags the end of a paragraph
$?	Controls echoing of subsequent records
$.	Flags the start of a topic

All tutorial actions are controlled with these four operators (along with the $* and $/ sentinels described above), making the task of creating a tutorial little more than the task of organizing a command file, documenting it with text, inserting control statements, and then extracting the flagged topics as a menu file (optional but recommended).

Tutorial Text: the $$ Flag. A line commencing with double dollar sign is by definition a comment, but it will be printed by TUTOR itself (simple $ comments are always passed on to the language system). TUTOR will strip off the $$ before printing the line. Furthermore, echoing of $$ comments cannot be suppressed while a tutorial is active, although the user can skip to the next topic by typing ".;".

Paragraph Demarcation: The $; Flag. A topic may consist of any number of paragraphs, each one of which is delivered to the user as a text block. At places where an author wishes to pause a tutorial (to let the user think and act on it) the sequence $; should be inserted as the first two characters in a record. The rest of the record will be printed, and control turned over to the user.

Controlling Echo: The $? Flag. Authors can suppress the printing of text and echoes of ODYSSEY commands in a tutorial file. This is mainly used to hide certain ODYSSEY commands from view during the course of a tutorial, as when initializing a program for special demos. This is done with the $? file command, which has four forms: $?0 (no echo); $?1 (echo input); $?2 (echo interpretation); $?3 (echo everything).

Tutorial Topics: The $. Flag. When a record is encountered by TUTOR which begins with the characters "$.", file reading is halted, the record is numerically indexed as a topic, and its text printed as a title. Input is then read from the user's terminal. The topic's index number is automatically derived from the number of periods (from one to ten) in its flag. A "top-level" topic is flagged by the two characters "$.". A second-level topic is denoted by "$..", a third-level topic by "$...", and so on. TUTOR will generate an index number in the form I.J.K... where I, J and K are 1 or 2 digit integers, contextually computed to the correct cardinality. Automatic numbering relieves authors of numbering tutorial topics, and protects users from mistakes that might be made in the process.

CLOSING THE LOOP

To fully realize its interactive potential, ODYSSEY needs a capability which is as-yet unimplemented. This might be called the capability to create command files (C^3F). Currently, one experiments interactively with parameters for maps and processing, toiling until the desired result is attained. To document the procedure arrived at, its inventor must remember all its details long enough to exit from ODYSSEY and type the commands into a text file. Doing this can be error-prone and tedious, and could easily be obviated.

All that would be needed to achieve C^3F is a procedure for recording commands typed into ODYSSEY, saving them as tutorial command files. The

TUTOR module could be enhanced to perform this without adding any new sentences to the command language, by intercepting users' comments. One way is to scan each comment typed in for one such as:

$#<filename>

This would signal TUTOR to create the file specified and begin echoing all user input into it. When the comment "$/" was typed, the file would be closed, just as tutorials are when being read. Other comments in the form of tutorial file commands can be entered, of course, or can be incorporated later by text-editing the new file, which will probably need to be revised in order to remove erroneous or superfluous commands.

ODYSSEY's language system is a powerful tool for interactive control of applications programs. Its many features, only some of which were discussed here, provide not merely a "user-friendly" environment, but one which is self-descriptive and extensible as well. Throughout the nine program modules, ODYSSEY commands follow uniform rules of syntax and employ a consistent, colloquial vocabulary. The system's tutorial facility backs up the command language, making documentation, advice and illustrations available to users during interactive sessions. More fluent users can add their own tutorials on any topic, extending the repertoire of conversations with ODYSSEY.

BIBLIOGRAPHY

<The following works offer the most useful introductions to the design and operation of ODYSSEY and its user interface.>

Broekhuysen, M., 1983. "ODYSSEY Geographic Information System Digitizing Guide." Cambridge, MA: Laboratory for Computer Graphics and Spatial Analysis.

Chrisman, N., 1979. "A Many Dimensional Projection of ODYSSEY." Cambridge, MA: Laboratory for Computer Graphics and Spatial Analysis.

Dougenik, J., 1978. "LINGUIST: A Table-driven Language Module for ODYSSEY." In _Harvard Papers on Geographic Information Systems_, edited by Geoffrey Dutton, Volume 7. Reading, MA: Addison-Wesley.

Dutton, G., 1978. "Navigating ODYSSEY." In _Harvard Papers on Geographic Information Systems_, edited by Geoffrey Dutton, Volume 2. Reading, MA: Addison-Wesley.

Laboratory for Computer Graphics and Spatial Analysis, 1983. "ODYSSEY System Summary." Cambridge, MA.

Morehouse, S. and Broekhuysen, M., 1982. _ODYSSEY User's Manual_. Cambridge, MA: Laboratory for Computer Graphics and Spatial Analysis.

White, A.M.H., 1980. "Creating Cartographic Data Bases in ODYSSEY." Cambridge, MA: Laboratory for Computer Graphics and Spatial Analysis.

White, D., 1979. "ODYSSEY Design Structure." In _Harvard Library of Computer Graphics_, Volume 2. Cambridge, MA: Laboratory for Computer Graphics and Spatial Analysis.

UN LOGICIEL DE CARTOGRAPHIE ASSISTEE PAR MICRO-ORDINATEUR

Jean-Paul Donnay
Université de Liège
Place du XX août 7
B4000 Liège Belgique

RESUME

Cette communication présente un logiciel original permettant la création de cartes thématiques sur une configuration centrée sur un micro-ordinateur. Quatre types de documents thématiques sont possibles - hachurage, formes proportionnelles, isolignes, 3-D - tandis qu'une base de données géographiques permet d'effectuer toute une série de traitements préalables. A cela se joint un ensemble de modules de traitements graphiques traditionnels qui permettent d'améliorer ou de compléter le dessin final. Le logiciel est totalement interactif. Sa structure modulaire lui permet d'évoluer aisément par ajouts de routines nouvelles. Traduit en deux langages - Basic compilé et Pascal sous CP/M - le logiciel peut tourner sur plusieurs systèmes. De même le software de commande des périphériques graphiques est court-circuité au profit de routines originales intégrées au logiciel de manière à rendre l'ensemble adaptable au plus grand nombre de configurations.

INTRODUCTION

Les logiciels de cartographie assistée par ordinateur sont le plus fréquemment liés à un hardware imposant (Boyle 1980, Nagy et Wagle 1979). D'une part la masse d'informations à saisir et d'autre part le degré de finition demandé par les agences publiques, généralement à l'origine de ces recherches, imposent en effet un équipement sophistiqué tant au niveau du processeur et des capacités de mémoire qu'au niveau des périphériques graphiques. Cependant, parallèlement à cet essor de l'infographie, l'apparition et la rapide banalisation du micro-ordinateur ont apporté un outil de travail plus à la mesure des petits centres de recherche tels nombre de services universitaires ou de bureaux d'étude privés (Toong et Gupta 1983). Moyennant un équipement minimum, cette génération de micro-ordinateurs est prête à accomplir un travail, sans doute peu en rapport avec les exigences d'une carte topographique, mais très efficace en matière de cartographie thématique. Diverses réalisations du Groupe Image de l'Université de Rouen notamment sont là pour le prouver (Leduc 1979).
Après s'être essayé à plusieurs applications cartographiques s'appuyant sur un système informatique relativement puissant (Donnay 1981), nous avons imaginé de réaliser un logiciel infographique intégré adapté au micro-ordinateur. Infographique car, à terme, ce logiciel doit être en mesure d'exploiter les routines graphiques de base à d'autres fins que la cartographie. Intégré car plusieurs programmes de traitement doivent être disponibles d'une manière transparente pour l'utilisateur et l'ensemble des opérations doit être prévu depuis la saisie de données jusqu'à la réalisation du dessin.

UN SYSTEME PORTABLE ET OUVERT

Dans la mesure où le logiciel est susceptible d'être utilisé dans divers environnements, il est souhaitable de le rendre indépendant de la machine sur laquelle il a été créé. De plus il est préférable qu'il présente une structure qui permette un enrichissement ultérieur éventuel. Ce sont ces deux notions que nous traduisons par portabilité et ouverture du système.

Un système portable

Rendre le logiciel indépendant de l'ordinateur est en première approximation une question de langage de programmation. Le rendre indépendant des périphériques graphiques utilisés demande une structure de programme adéquate et une nouvelle définition des tâches graphiques élémentaires.
En ce qui concerne le langage de programmation, et pour rencontrer une large gamme d'utilisateurs, une version du logiciel est écrite en BASIC étendu APPLESOFT - ordinateur utilisé : Apple II Plus de 48 K de mémoire centrale. Au moment de l'exécution la version interprêtée est abandonnée au profit d'une version compilée. La structure du logiciel nous a d'ailleurs amenés à utiliser les extensions de langage offertes par le compilateur TASC de Microsoft, notamment en ce qui concerne la réservation de variables communes (common).
Pour atteindre une portabilité libérée du matériel Apple, une version plus riche mais non encore totalement achevée est opérationnelle en Pascal/Z sous CP/M. Structure et place-mémoire y gagnent évidemment beaucoup.
Les tables traçantes offertes sur le marché diffèrent par leur degré d'intelligence, en d'autres termes par leur relative richesse en routines graphiques implantées et par l'accès qui leur est accordé. La diversité est particulièrement grande dans les bas de gammes équipant les micro-ordinateurs puisque certains traceurs sont fournis sans software intégré.
Pour permettre la portabilité du logiciel au niveau de ce périphérique, nous avons été amenés à créer un nouveau logiciel graphique de base, partie intégrante du logiciel général. Par ce fait, toute librairie éventuelle propre au traceur est court-circuitée. Les routines nécessaires aux applications sont ainsi fournies dans le langage et les conditions voulues.
En amont du logiciel de dessin lié à la table traçante, toute une série de modules de saisie et de traitement de données a été développée. Elle constitue un base de données réduite et fait référence à un autre périphérique : le digitaliseur. Il ne nous a pas été permis jusqu'ici de confronter les divers équipements de prise de coordonnées susceptibles d'être connectés online sur une configuration micro-informatique. Dès lors, jusqu'à plus ample informé, cette partie du logiciel reste liée au type de périphérique équipant notre installation, en l'occurrence une simple tablette graphique Apple.
Néanmoins les quelques instructions commandant l'enregistrement des couples (X,Y) sont confinées à une seule sous-routine aisément modifiable voire généralisable si la chose s'avère possible. Bien que cela ne soit pas imaginé comme solution alternative, signalons néanmoins que l'introduction de coordonnées peut être réalisée au départ du clavier.

Un sytème ouvert

La structure du logiciel doit permettre son enrichissement ultérieur éventuel. Dès le départ, c'est ainsi qu'il a été conçu. Chaque module de programmes ou sous-programmes du logiciel a été mis au point de façon indépendante puis inclus au système. Un structure hiérarchisée de menus permet d'activer isolément chacun des modules. Le logiciel graphique de base doit constituer une librairie de sous-routines accessibles depuis tous les modules. Néanmoins, en BASIC, une telle librairie ne peut être indépendante de son programme principal et il a fallu adjoindre à chaque module de tracé les sous-programmes graphiques qu'il est susceptible d'utiliser. Une redondance qui augmente sensiblement la taille des modules et qui est bien sûr évitée dans la version Pascal du logiciel.

STRUCTURE GENERALE DU LOGICIEL

Le logiciel est composé de deux ensembles de modules distincts : la base de données et le logiciel graphique proprement dit. Ce dernier est développé selon trois axes : cartographique, graphique et utilisation conversationnelle des routines graphiques de base.

La base de données

La première tâche de cet ensemble de modules est de constituer une banque de données géographiques dont dépendra toute application ultérieure. La technique de représentation a été déterminée sur base de trois composantes élémentaires : le noeud, la chaîne et la zone. Des relations de type hiérarchique entre ces trois éléments permettent la représentation de tout objet physique dans un plan. La technique est connue et largement discutée dans la littérature (Kobayashi 198Ø, Corona Burgeno 198Ø, Smedley et Aldred 198Ø). Rappelons que le noeud est un point situé au croisement d'au moins trois segments et qu'il est identifié par ses coordonnées planes. La chaîne est constituée d'une suite de segments - eux mêmes définis par les coordonnées de leurs points limites - entre deux noeuds. Elle porte également dans notre application les caractéristiques du tracé (épaisseur, continuité ou non, opérationnel ou non). La zone est définie par une suite de chaînes. La zone peut être ouverte (route) ou fermée (district) et ne comporter qu'une chaîne (zone fermée à une chaîne : île) ou plusieurs (zone ouverte à plusieurs chaînes : réseau de routes). Il va de soi que les exemples ne sont pas limitatifs.
La banque de données s'appuie ainsi sur quelques fichiers élémentaires à accès aléatoire et les interconnections sont réalisées au moyen d'une série de pointeurs (figure 1).
Un effort particulier a été consenti quant aux fonctions d'édition de la banque de données. Ainsi toute introduction ou définition de données élémentaires est soumise à une procédure interactive de type "full screen" et divers algorithmes de contrôle et de présentation standard ont été introduits (par exemple la présentation des chaînes constituant les zones selon le sens horaire). La structure de représentation choisie permet en outre une réalisation aisée des fonctions d'insertion, de correction ou de suppression d'informations élémentaires, notamment par le fait qu'elle interdit toute redondance de l'information.
La notion de base de données suppose l'existence de procédures de traitement de la banque de données, conduisant à une utilisation plus efficace de celle-ci. En effet, si la forme sous laquelle se présente la banque de données s'avère pratique pour l'édition et la mise à jour, l'intervention de multiples pointeurs hiérarchisés lors de la phase de lecture d'un élément complet (ainsi la frontière

d'une zone) est peu rentable lorsqu'elle se répète de façon continue (lors du hachurage de la zone par exemple).
Le premier rôle de la base de données est dès lors de fournir des fichiers distincts sous la forme la plus indiquée à leur utilisation ultérieure (essentiellement : limite en séquentiel, réseau quadrillé de points values, fichier texte et fichier de valeurs).
Il est utile de signaler que les nombreuses manipulations de fichiers susceptibles d'être réalisées par la base de données restent transparentes pour l'utilisateur sous couvert d'un nom générique d'application. Telle est en tout cas l'option par défaut.
En plus de la simple remise en ordre des informations, la base de données peut effectuer quelques traitements préliminaires. Les deux principaux actuellement disponibles sont la généralisation du tracé (optionnelle) lors d'une modification d'échelle et la concaténation d'éléments définis dans la banque de données (agrégation de zones en aires plus vastes de niveau hiérarchique supérieur). L'extraction de fenêtres constitue aussi une tâche importante qu'il est nécessaire d'inclure parmi les fonctions d'une base de données.
Dans le cas particulier des zones fermées, l'affectation de labels et leur positionnement automatique au sein de la zone constitue une des facilités offertes par la base de données. Il s'agit en fait de créer un fichier texte particulier et de modifier les caractéristiques de la frontière de la zone pour y inclure une enclave détourant le label en question (figure 2).
Plus simplement et sans détourage, le logiciel prévoit en outre l'affectation de labels aux noeuds du graphe.
Enfin les valeurs d'une variable peuvent être associées aux zones ou aux noeuds du sytème. Ces fichiers de valeurs représentent en fait le phénomène à cartographier.
Toute constitution de banque de données géographiques est longue et leur relative complexité nécessite une préparation minutieuse qui en augmente encore le coût. Malgré le peu d'attributs présents dans ce système - vis-à-vis des banques de données urbanistiques ou géologiques par exemple - son emploi requiert la consultation d'un protocole d'utilisation qui peut seul en définir toutes les possibilités et les contraintes.

Le logiciel graphique

Le coeur de ce logiciel est constitué d'une douzaine de sous-routines dites de base qui permettent de répondre aux spécifications des programmes plus évolués. Dans la version actuelle du logiciel, les principales routines sont les suivantes :
- Initialisation d'un espace virtuel de travail;
- Déplacement entre deux points par interpolation linéaire (épaisseur et continuité variables des traits);
- Définition d'un jeu de 64 caractères et symboles et positionnement de texte quelconque ou hiérarchisé autour d'un point ou d'une forme;
- Définition de formes d'orientation et de taille variables;
- Tracé d'axes gradués, cotés et commentés.

Lors de la mise en route du système, le logiciel passe par une phase d'initialisation. Celle-ci permet la lecture - ou la définition - des paramètres de configuration (caractéristiques du traceur), la définition des variables communes et, seule tâche réservée à l'utilisateur dans cette phase du travail, la définition d'un espace virtuel.
Ensuite apparaît le premier aiguillage appellé Menu Général. Il peut être activé soit par la procédure d'initialisation soit par les menus de second ordre en retour d'exécution (figure 3). Le menu

général permet de choisir entre les trois manières d'utiliser le logiciel graphique, à savoir : le module conversationnel, le module cartographique et le module graphique. Rappelons que la structure en arbre adoptée permet d'utiliser successivement ces différents modules lors d'une même application.
C'est précisément dans ce but qu'a été développé l'accès conversationnel aux routines de base. Le module conversationnel permet de définir d'une manière interactive les différents arguments nécessaires à l'exécution de chaque routine de base. De la sorte il est possible d'effectuer tout travail graphique selon une procédure pas-à-pas. Mais son utilité essentielle consiste à fournir un outil de finition d'un travail. Ajouter un commentaire, un cadre ou un symbole sont autant de réalisations qui peuvent être effectuées une à une en dehors ou en sus d'un programme d'application traditionnel. Les quelques graphiques illustrant ce texte sont réalisés de cette manière.
Le module cartographique prévoit quatre procédures de traitement accompagnées de deux programmes de mise en page. Ceux-ci réalisent d'une part le fond de carte et les écritures et d'autre part l'habillage des cartes (échelles, direction du nord, titre, cadre, commentaires divers). Ces deux programmes peuvent être appelés isolément ou en complément d'une des quatres procédures de traitement cartographique.
Ces dernières permettent le tracé de cartes choroplèthes, le positionnement de formes de tailles proportionnelles, le tracé d'isolignes et la perspective 3-D. Chacun de ces traitements a fait l'objet, antérieurement, d'un programme spécifique écrit en Fortran et installé sur une configuration plus importante. Le tracé d'isolignes et la perspective 3-D installés sur micro ne sont d'ailleurs que des traductions plus ou moins bien adaptées de ces programmes originaux. Toutes deux seront améliorées dans un proche avenir.
Les procédures de hachurage et de positionnement de formes sont par contre entièrement revues en fonction de leur utilisation sur micro-ordinateur. La première réalise le hachurage de zones suite à la lecture de valeurs préalablement classées. Les trames utilisées sont soit pointées dans une "banque de 24 trames" standards, soit définies en tous leurs paramètres par l'utilisateur.
La même affectation de trames peut être utilisée pour hachurer des formes centrées sur les noeuds de la carte et dont la taille est proportionnelle à la valeur de la variable à cartographier. Cinq formes de base sont prévues : rectangle, triangle équilatéral, hexagone, demi-cercle et cercle complet. Les formes asymétriques peuvent être orientées dans deux ou quatre directions. Lorsque la routine de hachurage est appelée, elle propose une légende reprenant les trames utilisées accompagnées de commentaires optionnels.
Le tracé d'isolignes suit la technique des moyennes mobiles à deux dimensions. Le semis de points valués constituant le fond de carte est lissé en un réseau quadrillé lors de la phase "base de données" (voir supra). La procédure de tracé réalise l'interpolation linéaire entre les sommets du réseau carré. Les cotes des isolignes sont indiquées mais jusqu'ici aucun algorithme de lissage des courbes n'a été introduit.
La procédure 3-D est liée au traitement d'isolignes. Le même réseau quadrillé sert de référence de départ et plusieurs routines sont communes. Moyennant quelques paramètres (angles de rabattement et de rotation, exagération des hauteurs,...) une vue classique de perspective axonométrique est dessinée sur traceur. Les lignes

cachées n'apparaissent pas et des commentaires sont possibles (échelles spécifiques par exemple).
Le module graphique constitue la dernière façon d'utiliser le logiciel. Cette dernière utilisation rassemble quelques programmes réalisant le dessin de diagrammes conventionnels. Les graphes de fonctions mathématiques demandent la définition d'un espace virtuel de travail et de la fonction à dessiner. Ils appellent ou peuvent appeler les routines de base de tracé d'axes, de positionnement de texte et les différents types de traits.
L'histogramme et le diagramme en quartiers de tarte, qui constituent les deux autres représentations graphiques implantées, peuvent en outre faire appel à la routine de hachurage et à sa légende optionnelle. D'autres diagrammes peuvent être introduits dans le logiciel sans aucune difficulté.

CONCLUSION

On a pu constater que le logiciel, tout en étant opérationnel, demande encore plusieurs améliorations (traitements d'isolignes et 3-D) ou compléments de fonctions (nouvelles fonctions de la base de données ou nouveaux diagrammes). En fait depuis sa création, le logiciel a connu de multiples mises à jour en fonction des demandes formulées par les utilisateurs. La structure ouverte qui lui a été conférée a toujours permis ces nouvelles implantations. Parmi les tâches qu'il est prévu de réaliser, on retiendra encore la connexion des modules graphiques et cartographiques (association de diagrammes à des éléments du fond de carte) et, à plus long terme, la traduction intégrale du logiciel en Pascal U.C.S.D. qui en permettra notamment une utilisation beaucoup plus souple sur Apple II, équipé seulement d'une carte langage supplémentaire.

REFERENCES

Boyle A.R. 198Ø, Development in Equipment and Techniques: The Computer in Contemporary Cartography (Taylor D.R.F. ed.): Progress in Contemporary Cartography, Vol. 1, pp. 38-57

Corona Burgeno J.F. 198Ø, A Geographical Data Base: Data Base Techniques for Pictorial Applications (Blaser A. ed.): Lecture Notes in Computer Science, No. 81, pp. 347-363

Donnay J.P. 1981, Cartographie par isarithmes. Présentation d'un programme utilisant la méthode des moyennes mobiles à deux dimensions: Bulletin de la Société Géographique de Liège, No. 16-17, pp.7-17

Kobayashi I. 198Ø, Cartographic Databases: Pictorial Information Systems (Chang S.K. et Fu K.S. ed.): Lecture Notes in Computer Science, No. 8Ø, pp.322-35Ø

Leduc A. 1979, Le système CARTOVEC: Cahiers Géographiques de Rouen, No. 1Ø-11, pp.1Ø3-133

Nagy G. et Wagle S. 1979, Geographic Data Processing: Computing Survey, Vol. 11, pp.139-181

Smedley B. et Aldred B. 198Ø, Problems with Geo-Data: Data Base Techniques for Pictorial Applications (Blaser A. ed.): Lecture

Notes in Computer Science, No. 81, pp.539-554

Toong H.D. et Gupta A. 1983, Les ordinateurs individuels: Pour la Science (Scientific American), No. 64, pp.46-61

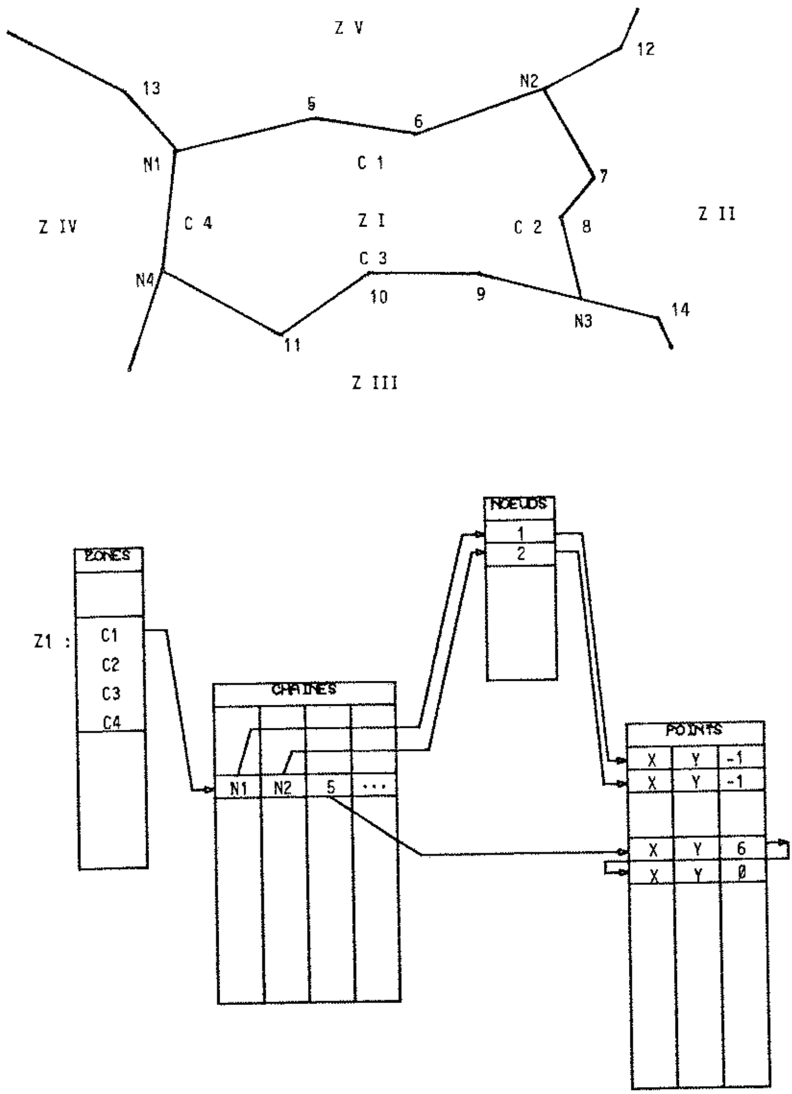


FIGURE 1. - BANQUE DE DONNEES
FICHIERS ELEMENTAIRES ET POINTEURS.

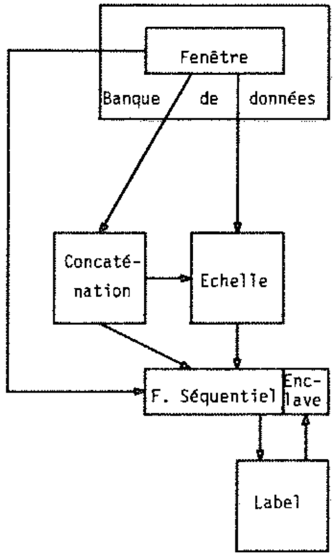


FIGURE 2. - BASE DE DONNEES
EXEMPLE DE TRAITEMENT EFFECTUE
SUR LA BANQUE DE DONNEES.

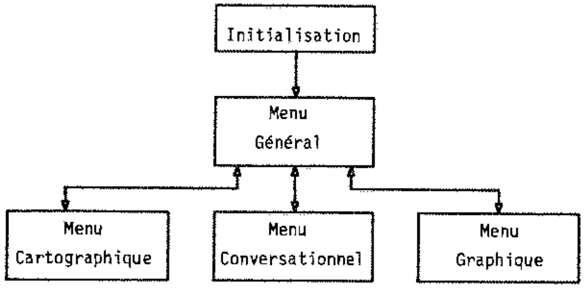


FIGURE 3. - STRUCTURE HIERARCHIQUE
DU LOGICIEL GRAPHIQUE.

LA CAD
A COMPUTER-ASSISTED DESIGN PACKAGE FOR MICRO-COMPUTERS

Robert M. Itami
Randy Gimblett
School of Landscape Architecture
University of Guelph
Guelph, Ontario N1G 2W1

ABSTRACT

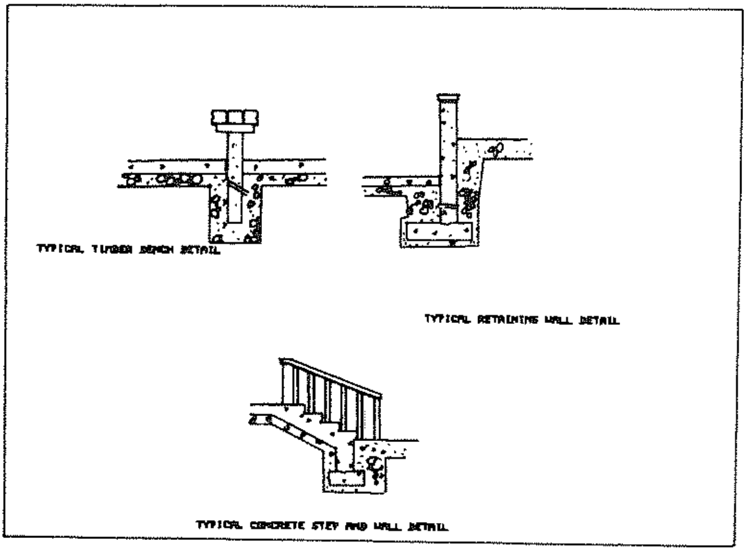
With the rapid decrease in cost of computer graphics hardware, it is rapidly becoming feasible for small design offices to take advantage of the benefits of computer assisted design and drafting packages. This paper describes a software package called LA CAD developed for applications in Landscape Architecture. The package allows interactive graphic input of maps, base plans and sections, storing them in sequential access disk files. It provides a facility for overlaying planting plans onto base plans and interactively simulating the growth of plants from any user-selected time in the future.

The program draws on user-defined graphic libraries of standard plant symbols, construction details and other graphic symbols in plan or section and scales them to the drawing. Output programs allow the user the option to select the scale of the output, edit and update graphic files, and manipulate the location of graphic objects on the screen. LA CAD provides some powerful capabilities for computer assisted design for use on low-cost micro-computers

INTRODUCTION

Computer assisted design and drafting packages have been available for some time on large, expensive mini-computers and main-frame computers. Because of the high cost of these systems the advantages of automated design and drafting have remained out of reach of small design firms and schools. With rapid development of micro-computer technology and the concurrent reduction in their cost, many of the features of large CAD systems may be provided through low-cost micro-computer technology. For the small office, a micro-computer based drafting and design system must not only be relatively low cost but must also perform routine tasks with speed, ease of use, and accuracy. In addition such a system should take advantage of the more powerful potentials of the computer which provide the user with capabilities which may have been too expensive to execute manually.

With these thoughts in mind the authors have developed a computer assisted design package called Landscape Architectural Computer Assisted Design (LA CAD). LA CAD performs most routine drafting tasks plus some capabilities



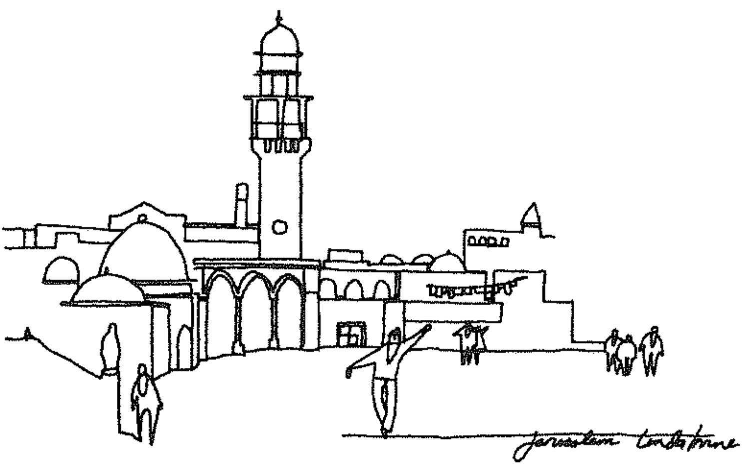


FIGURE 1: Sample output from LA CAD. Top illustration shows plotted output of construction detail selected, scaled and plotted from user-defined library. Bottom illustration shows sample output from MDRAW.

designed specifically for landscape architectural design applications. Briefly, LA CAD provides the following functions:

1) Interactive input and editing of drawings (base plans or sections) incorporating text and graphics.

2) Creation of user-defined graphic libraries.

3) Creation of planting plans or sections using standardized symbols from plant libraries.

4) Simulation of plant growth at any user-defined period of time in the future.

5) User-controlled selection, location and scaling of standard symbols from user-defined graphic libraries onto base sheets.

6) Outputs to color graphics terminal or digital pen plotter with user controlled scaling.

These capabilities are discussed in more detail in the remainder of this paper.

APPLICATION OF LA CAD FOR DESIGN

In a typical design, the Architect and Landscape Architect will draft up a base plan which is used as a base for subsequent sketch plans and refinements. This function is provided by LA CAD using the program module called MDRAW.

This module allows the user to create a new graphic file or update an existing graphic file. Using a keyboard and a digitizer the user may select the color of the plot, and whether to draw vectors, polygons, circles, filled circles or include text. The program gives simultaneous graphic feedback on a color monitor during the digitizing process. The user may then check the input for accuracy before saving the work in the file. This process significantly reduces user errors and an inexperienced user may produce an error-free file in his first session. (See figure sample output figure 1.)

As the design is refined, the base plan will usually remain unchanged except for the addition of standardized symbols or additional graphic data. LA CAD eliminates the necessity of re-digitizing the same base plan by providing a means of graphically overlaying drawings created and saved in different files.

In the case of a planting design, the Landscape Architect will normally use only a limited number of plant symbols to illustrate the planting plan. LA CAD provides a facility for performing this operation in the program module referred to as LAYOUT. (See figure 2.)

LAYOUT allows the user to select any baseplan or section file created by MDRAW and create a separate design file

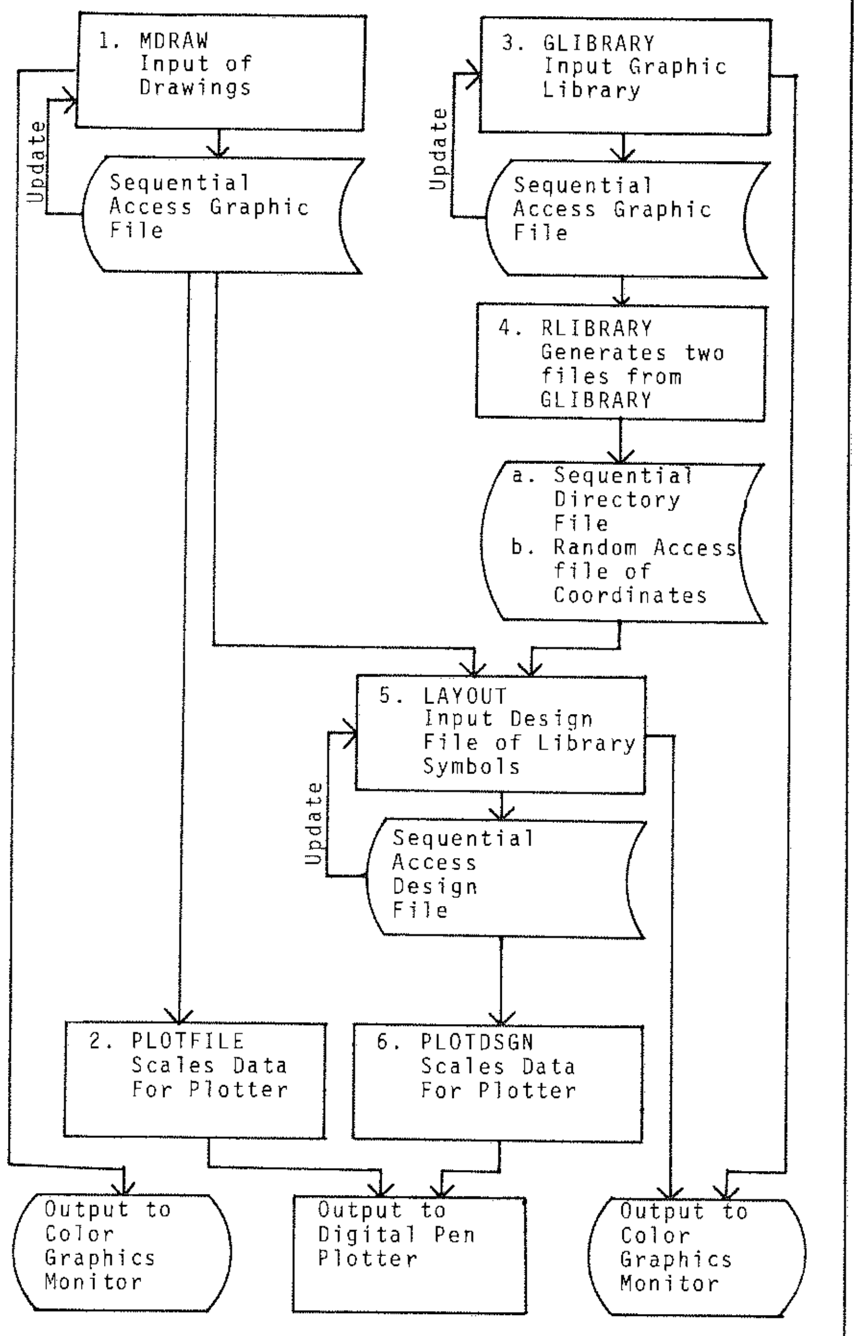


Figure 2. FLOWCHART OF LACAD Program Modules

which is graphically overlayed onto the base. Like MDRAW, LAYOUT allows the user to create a new design file, or update an existing file. Because the design files are handled separately from the base plan, it is possible to create alternative planting designs or other drawings for the same base plan without having to re-digitize or make duplicate copies of the base plan file.

LAYOUT accesses preconstructed graphic libraries of plant symbols, construction details or other user-defined libraries. The user may then choose the desired symbol from the library and scale and locate the symbol on a base plan. This significantly reduces the time normally required to create planting designs and provides consistently high quality graphics even for preliminary concept plans. (See figures 3, 4.)

LAYOUT provides two different facilities for determining the scale of symbols to be plotted on the base sheet. In the case of graphic objects such as construction details, the user may select the exact ratio scale at which s/he wishes symbols to be plotted. In the case of planting designs, LAYOUT offers some dynamic capabilities for simulating the growth of plants.

Plant libraries store not only the coordinates and names for plant symbols, but also the age of maturity for each plant. The symbols are digitized at mature size. When LAYOUT accesses these plant symbol libraries the user selects an age for the plant at planting time. The program then reads the coordinates from the file and scales them according to the percent age of the plant (age at planting/age at maturity), and the scale of the drawing (scale of plant library/scale of drawing). The absolute coordinates of the plant symbol are converted to relative coordinates and the plant is plotted on the color graphics terminal. (See figure 4.)

The user may mix symbols from different libraries in the same design file. When the design is completed, the user may redisplay the planting design at any time in the future.

LAYOUT displays the file directory and requests that the user select the appropriate file. It then displays the base plan on the graphics terminal. The program then pauses and displays the following prompt:

> THIS PROGRAM ALLOWS YOU TO VIEW THIS PLANTING DESIGN AT ANY TIME IN THE FUTURE FROM THE TIME OF PLANTING. PLEASE ENTER THE NUMBER OF YEARS IN THE FUTURE THAT YOU WISH TO VIEW THIS DESIGN.

The user inputs a numeric response and the program proceeds to read the plant design file scaling the plants according to the year selected. After the planting file is plotted, the program allows the user either to exit the program or add to the file. If the design file includes symbols other than plants, LAYOUT scales these according to the ratio scale requested by the user in the original input sequence.

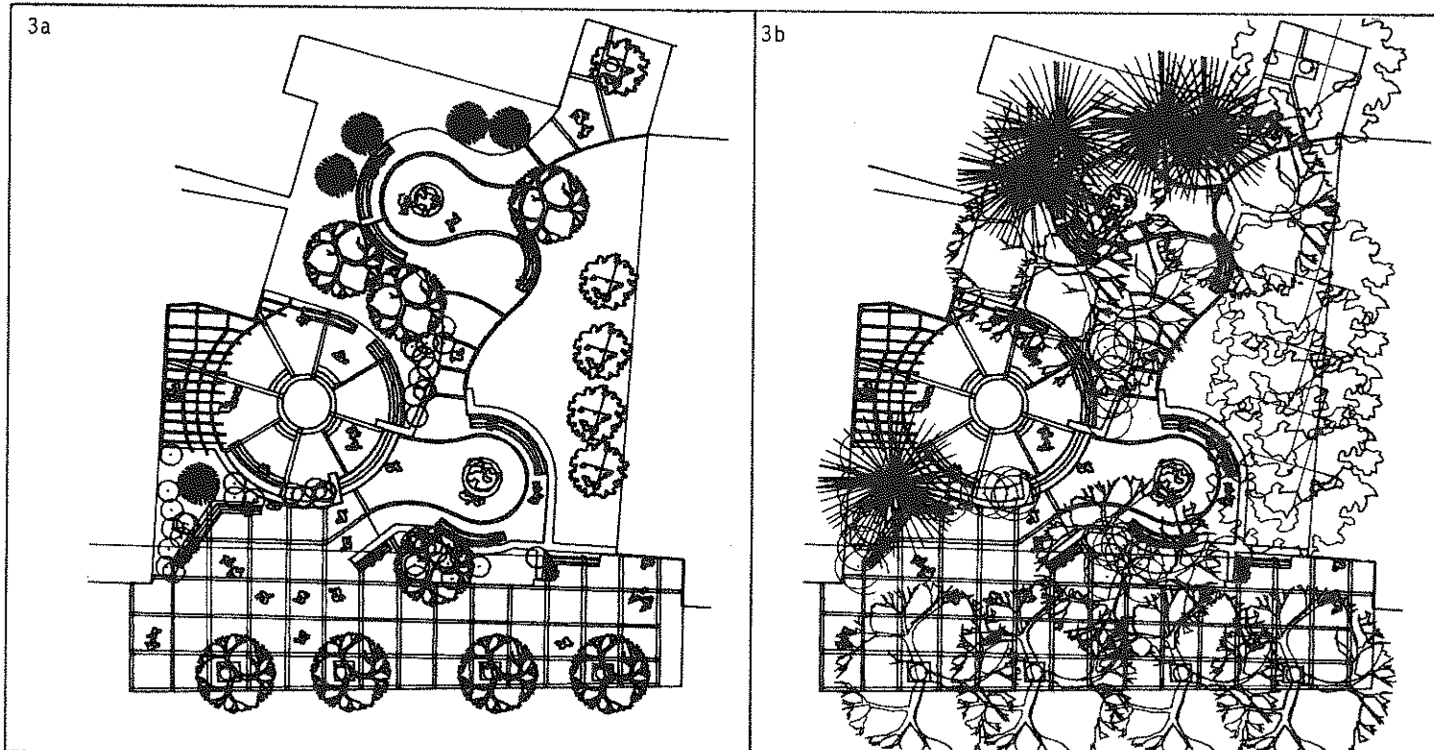


FIGURE 3: PLANTING PLAN FOR URBAN PARK DESIGN. Base plan digitized at 1:75 scale using MDRAW. Planting plan overlay created with LAYOUT. 3a and 3B plotted with PLOTOSGN at 1:200. 3a shown at time of planting, 3b twenty years later.

Once design files are completed to the satisfaction of the user, LA CAD provides two different program modules for producing paper copy on a digital pen plotter. If the user wishes to plot only the base plan created with MDRAW, the program called PLOTFILE is executed. This program allows the user to select the scale of output before plotting.

To plot the design files created by LAYOUT the program PLOTDSGN is used. This program is analogous to PLOTFILE. PLOTDSGN reads the file created by LAYOUT and outputs to the digital plotter. It first allows the user to select the scale to plot and then plots the base plan. After completing the base plan, the program asks the user to input the number of years in the future from the time of planting to display the plants. Then the program scales and plots the symbol's coordinates in the same manner described above for LAYOUT.

PROGRAM STRUCTURE OF LA CAD

The flow chart in figure 2 illustrates the program modules of LA CAD. Each of these modules is described below:

1. MDRAW - Interactive input and editing of base plans or sections. Outputs a sequential access file for the drawing.

2. PLOTFILE - Reads files created by MDRAW. Scales data to a user-defined scale and outputs to a digital plotter.

3. GLIBRARY - Input of user defined graphic libraries for planting design symbols and other standard graphics such as construction details. Outputs to a sequential access file.

4. RLIBRARY - Reads the sequential access file created by GLIBRARY and outputs two files: 1) a sequential access directory for graphic symbols, and 2) a random access file of x-y coordinates for the symbols.

5. LAYOUT - Interactive input and editing of design files for any user-defined base plan or section. Displays plant symbols or other graphic symbols from graphic libraries immediately upon selection on a color graphics monitor at the age and location designated by the user. Outputs a sequential access file.

6. PLOTDSGN - Reads files created by LAYOUT. Outputs graphic symbols and simulates plant growth with output to a digital plotter.

FEATURES OF LA CAD

<u>User Defined Libraries</u>. In order to reduce the storage requirements for design drawings and provide flexibility and speed in defining and plotting standard symbols, LA CAD provides the user with a means of creating graphic libraries.

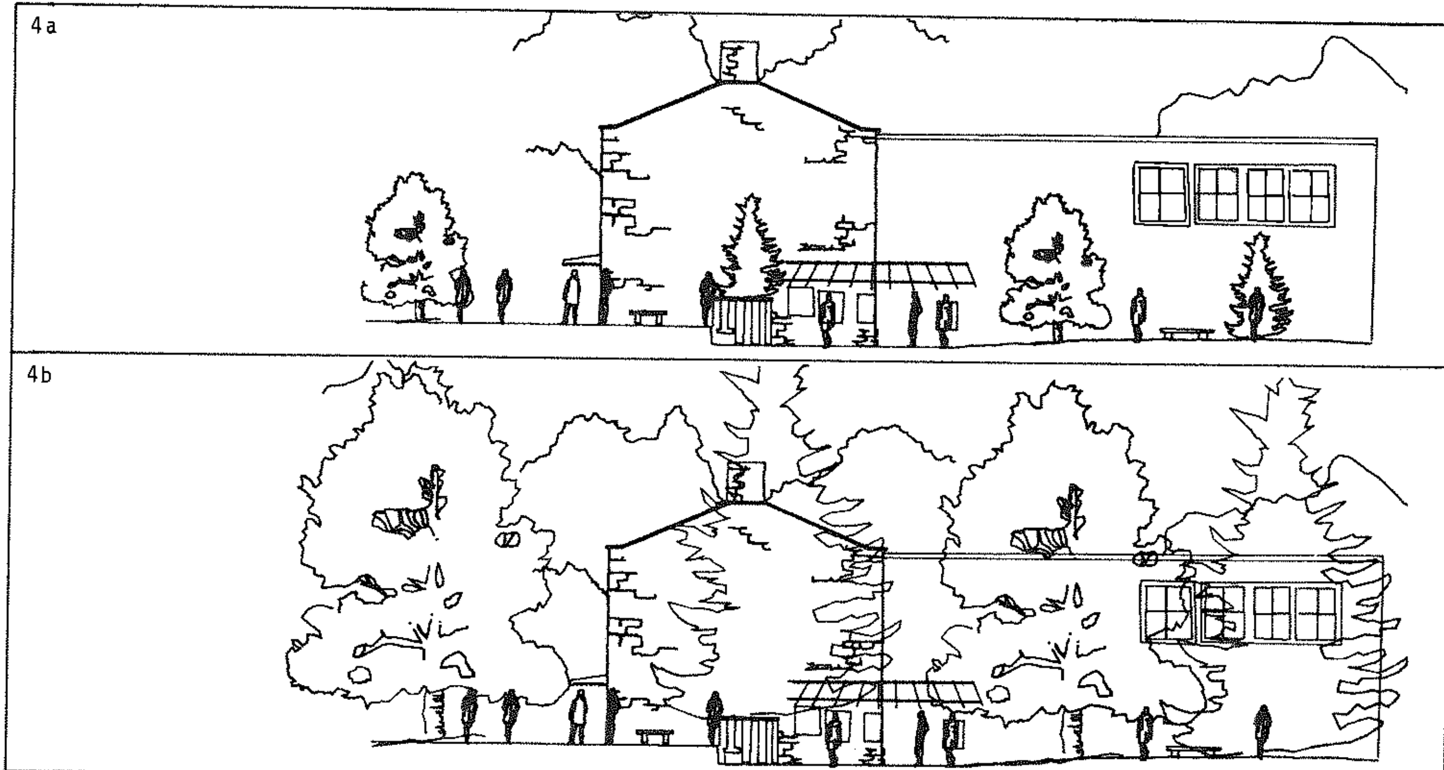


FIGURE 4: SECTION FOR URBAN PARK DESIGN. Original digitized at 1:50 using MDRAW, this example plotted at 1:200 using PLOTDSGN. All plants and people are drawn from libraries created using GLIBRARY and RLIBRARY. 4a shows design at planting time. 4b twenty years later.

GLIBRARY allows the user to create symbols of multiple colors or shapes. The symbols may therefore be a composite of smaller shapes or symbols. Graphic libraries may be updated and expanded with this program. Because of the length of time it takes to search sequential access files for symbols, another program called RLIBRARY reads the sequential access file created by GLIBRARY and creates two new files. The first is a random access file which stores the coordinates of graphic symbols in binary form. The second is a sequential access directory file which lists the name of the symbol, its scale and the record number of the coordinates for the symbol in the random access file. This structure speeds plotting time and reduces disk storage space.

Menu Driven Programming. LA CAD is completely menu driven, making it easy and simple for the user to learn and use. The menus are arranged in a hierarchy so the user has complete information at the appropriate level in the decision making structure.

Error Checking. LA CAD checks all user inputs for null answers, appropriateness of string responses or correct ranges for numeric responses. If the user fails to provide an appropriate response, additional prompts are included to assist the user before inputting again. LA CAD also checks for transmission errors which result from loss of coordinates, and for erroneous points digitized outside the boundaries of the drawing. In addition, the user has a continuous visual check of the digitizing since there is simultaneous output on a graphics terminal. This results in a high percentage of error-free files.

Scaling. LA CAD gives the user complete control of the scale of output to the plotter.

Multiple Overlays. The user may wish to overlay more than one drawing. For instance, he may have a base plan upon which he would like to overlay a file of dimensions for a layout plan, or a file of shadows and site furnishings for an illustrative plan. In order to accomplish this he may merge the files using a word processor, or he may overlay the drawings on output by plotting one file over the other.

APPLICATIONS

Education

LA CAD was developed in an educational environment to explore the use of micro-computer technology in Landscape Architecture. The concept of simulating plant growth over time is a powerful teaching tool. The student has the advantage of quick and easy feedback to design decisions. The simulation process allows the student to ask "what if"? questions about his/her design and quickly receive answers. Using this process, high quality planting plans may be easily created and alternative ideas explored. When manual techniques are used, however, a student must spend hours on the graphics of an initial design scheme - s/he often has neither the time nor the energy left to explore alternative designs. Essentially the student decreases the time dealing with manual techniques and can concentrate on the creative process in a dynamic and stimulating fashion. LA CAD serves in the educational environment well.

With future enhancements of LA CAD listed later in this paper, the student will have other tools to work with which will allow for the integration of a number of design considerations, including cost and site engineering factors in the design process.

Professional Practice

Plans are being made for testing LA CAD in the setting of a small professional office. It offers some promising opportunities for the small office to utilize low-cost micro-computer technology for computer assisted drafting and design. With the ability to create user-defined graphic libraries, the office can develop and maintain a "personality" in its graphics. The real test of the viability of a pcakage such as LA CAD in a professional environment is ease of use, time savings and quality of output. LA CAD fulfills these measures admirably. It generally requires only one hour of training for the average person to use LA CAD with confidence. The experienced user can generate drawings in favorable time compared to producing them by hand. The user then has the advantage of easily changing scales and producing quality "originals" at will. The time usually needed to create illustration quality planting plans is greatly reduced with the use of plant libraries, and even conceptual planting plans may be easily produced in the quality of a final drawing. With the rapid decrease in cost of graphic peripheral devices such as plotters, digitizers and graphic terminals, packages such as LA CAD show promise in providing low cost, computer assisted design services "in house" for the small design office.

FUTURE ENHANCEMENTS

The following features are planned for enhancement of LA CAD:

1) Improvement of algorithms for simulating plant growth. The present program simulates growth based on percent of mature age. This linear projection is not accurate for most plants since many plants grow fastest in youth. This component is not easily improved because of the lack of long-term data on plant growth under verying environmental conditions. (See reference list.)

2) A utility for creating libraries of different lettering styles (e.g. hand lettering) for fast "personalized" lettering of drawings.

3) The creation of graphic "data bases" for construction details and standard symbols (such as logos).

4) Cost estimating for planting designs. The planting design file created by LAYOUT can be read using another

program and compared to a current cost file of plants to produce automatic cost estimating for planting plans.

5) Site engineering utilities for calculating cut and fill, area, length, horizontal and vertical alignment, etc.

6) Adaptations are currently being studied for using LA CAD as a tool in computer assisted education. These experiments will explore the further development of the simulation capabilities of LA CAD and add programmed learning modules to the basic data structure of LA CAD.

PROGRAM SPECIFICATIONS

Name: LA CAD
Author: Robert M. Itami
Randy Gimblett
School of Landscape Architecture
University of Guelph
Guelph, Ontario, Canada N1G 2W1
Date: June 1983
Language: Micro-soft Basic-80
Operating System: CP/M or MP/M II (Digital Research)
Hardware Requirements:
Current Implementation

CPU	Ithaca Intersystems 525 c/ MP/M II operating system Minimum 64K Ram per user 5 meg hard disk
Digitizer	Calcomp Series 9000, 36" X 48" tablet
Graphics Terminal	Tektronix 4027a
Plotter	Calcomp 81, 8-pen plotter

Software Modules:

MDRAW	18K, 375 lines with comments
PLOTDSGN	14K, 325 " " "
PLOTFILE	10K, 208 " " "
LAYOUT	16K, 346 " " "
GLIBRARY	18K, 388 " " "
RLIBRARY	4K, 87 " " "

Parts of this paper are revisions of an earlier paper presented at the 1983 Computer Graphics Conference at Harvard University Graduate School of Design entitled "LA CAD: A Computer Assisted Planting Design Program for Micro-Computers" by Robert M. Itami.

REFERENCES

Collins and Moon Ltd. 1982, CAPE: Computer Assisted Plant Evaluation. A software package using D-Base II by Ashton Tate. Guelph, Ontario.

Finkel, L. and J.R. Brown. 1981, Data File Programming in Basic. John Wiley and Sons, Inc., New York.

SMALL AREA MAPPING SYSTEM (SAM)*

Shiu-Yeu Li, Department of Regional Industrial Expansion,
Ottawa, Canada, K1A 0M4
John Morrison, NORDCO, St. John's, Canada, A1B 3T2

ABSTRACT

The Small Area Mapping System is designed to meet the information requirements of decision makers responsible for formulating regional industrial policies and programmes in Canada. In the first part of this paper, the institutional environment in which the system was conceived, developed and implemented is described. The potential role of this system in the policy making process is illustrated, followed by highlights of its development history, with particular emphasis on the importance of interaction between development staff and users. The next section focuses on the characteristics of the system, containing a detailed description of User Interface, Data Linkages and Geograhic Data Base components. This is followed by an overview of system capabilities. The paper concludes with an outline of possible future enhancements.

INSTITUTIONAL ENVIRONMENT

Background

The mandate of the Department of Regional Industrial Expansion is to deliver programs to assist industries in all regions of Canada. Towards this end, the department is committed to develop regionally sensitive programs and policies, which seek to direct additional federal assistance to areas which are economically disadvantaged.

To assist policy and program formulation, it is imperative that relevant, timely and accurate information be available for a range of geographical areas. This includes provinces, regions and sub-provincial areas such as counties and communities. On the basis of these data, the decision-maker can assess the socio-economic problems of specific areas, formulate appropriate policy or programs, and evaluate their impact and effectiveness in tackling cyclical and structural problems.

An important system requirement is rapid response time. Issues emerge quickly and as new policy initiatives are identified, the options must be evaluated within strict time constraints, in order that timely and effective governmental actions be taken. Therefore, it is seldom feasible to develop a statistical data base which will be useful only for a single policy decision. Instead, all frequently requested data must be assembled and stored in a readily accessible manner.

---

* The views expressed in this paper do not necessarily represent the official positions of the respective organizations to which the authors belong.

Another major requirement of this information system is the ability to conceptually illustrate the intended message of statistical data. Graphic presentation is the most effective method for quick, easy and above all accurate assessment of a given situation. Far superior to the summary of numerical information in tables, the use of maps permit better understanding of the spatial distributions of variables and the relationships between them.

The maps produced by this system are not really maps in the sense of topographic maps, rather, they are cartograms - a grahic form which provides both a sense of geographic location and some statistical information. Precise location and complete accuracy of the statistics are not important. What is important is that the visual information convey the intended information. It should be noted that the majority of the users do not have cartographic training and are not familiar with these types of maps. Therefore, depending on the specifications of management, certain conventional cartographic principles may have been over looked in producing some of our maps.

Role of SAM in Policy Analysis Process

SAM has proven to be an extremely useful tool in assisting decision makers to appreicate a given situation. An example of a map produced by the system is found in Appendix I. Typically, the role of SAM in the process of policy making would consist of the following simplified steps. It should be noted that this is an iterative process. The actions taken in progressive steps follow directly from the results of the previous action. The difference between this process and those in the academic world is that, instead of incorporating inputs from peers, the decisions of government officials shape the action to be taken at each step.

1. Conception of problem - originates from any of the Minister (MIN), Deputy Minister (DM), Assistant Deputy Minister (ADM), directors or analysts.

2. Design of study - directors, analysts.

3. Collection of data - analysts, support staff.

4. Preparation of maps - support staff.

5. Formulation of policy recommendation and review of map by analysts, if modification is required, go back to step 2, 3 or 4.

6. Similar review process by directors, ADMs, DM and MIN, go back to steps 2, 3 and 4 if necessary.

7. Similar review process by the Committee of DMs, Committee of MINs and finally by the Cabinet, go back to step 2, 3, and 4 if necessary.

8. Policy recommendation considered by the cabinet.

9. Legislation passed in the House of Commons and Senate.

Depending on the complexity of issues, the process from the conception of problem to the legislation stage is a long one, involving numerous modifications to the maps. It is, therefore, extremely important that maps be produced very quickly, without unnecessary attention to "perfection", and be understood readily by individuals without training in cartography. SAM satisfactorily fulfills these requirements.

DEVELOPMENT HISTORY

The Small Area Mapping System has a long history of evolutionary development. It began, in fact, as two separate systems, one called MAPZ which was used to plot county data and the other COMPLOT which was used to plot communities. These initial systems, while received with great enthusiasm by the user community, were very primitive both in terms of their cartographic capabilitites and their operational characteristics. Unfortunately, the enthusiasm did not extend to the allocation of funds for the purchase of either new software or hardware.

The first revision to the systems was conducted in our spare time between peaks of activity in our 'real' work. This first effort was primarily directed at increasing the operational efficiency of the system. In its first incarnation, the system could only be run by a programmer and was both slow and very inefficient.

Three key decisions were made in this spare time work on the proto-SAM system:

1. The use of a cartographic database philosophy was decided upon.

2. The COMPLOT system was incorporated into MAPZ. Having made this decision at this time, it followed that whenever a new geographic data base was required, it would be integrated into one system rather than be set up as a parallel system.

3. The philosophy of user interaction was established. Our goal was that maps could be produced by people who had neither cartographic nor computer background, after a single 20 minute demonstration.

As this first major revision of the mapping system neared completion, an experience occured which was to be repeated in every subsequent release of the system. There was an immediate requirement for a series of maps that had features only available in the as yet untested new system. In order to produce these maps, the programmers had to work closely with the analysts and support staff in creating production maps from a very creaky system. This seemingly innocuous event affected all of the subsequent development for the next two years.

Our programmers began to see, many of them for the first time, their work in a production environment. They saw how a design feature, that had little consequence to the programs, could create tremendous problems in the operation of the system. They experienced how frustrating it was to work for many hours after the regular quitting time and watch a slow program perform. They learned the exasperation of discovering an irrecoverable mistake, where the whole procedure had to be restarted from the very beginning. Most important of all, they were on hand when the users asked often naive but sometimes very penetrating questions about why the computer did things "that way".

After that first experience, every effort was made to see that the development staff and the end users shared the same facilities. The payoff has been tremendous in terms of subsequent system design. The benefits of cross fertilization of ideas between user and programmers cannot be overemphasized. The systems have become easier to use, more closely tuned to the requirements, and much more reliable. Not only did the development people learn about the user environment, but users gained a better understanding of the problems and capabilities of the developers. As a result, requests for development also improved in quality.

Once the major design parameters had been established, and an understanding between the development staff and the users was established, remaining work progressed smoothly. As of today, five versions of SAM have been released to users across Canada. Each release has added new features to the maps, has been more efficient and has been easier to use.

SYSTEM CHARACTERISTICS

The most important features of the SAM system are the User Interface, Data linkages and Geographic DataBase components. The underlying principle in the design of each of these components was to meet the requirements elaborated in the previous sections.

User Interface

The User Interface has been designed to facilitate high throughput and flexibility. In order to achieve both of these somewhat conflicting objectives, we designed the system using the block mode full screen interface concept. The various parameters that are required to define a map are presented to the user in logically related groups on the screen. The user then enters all the values for one aspect of the map at one time. This process is as simple as filling in blanks on a form.

In all cases, the system displays the default values for each parameter. Herein lies the real strength of the system. The default values are stored in a file that is read by the program for the start of each new map. The user tells the system which file to be used for the defaults. At the end of each map definition, the user can save the parameters in the parameter file, and they can then be used as the default parameters for the next map.

The control system is set up such that a plot can be requested at any time using the special function keys. This means that to reproduce a map, one simply designates it as the default and then requests a plot without changing any of the parameters. For a series of maps, a default is selected and then only the few parameters necessary to produce the series need be set for each new map in the sequence.

Other function keys permit a rapid scan of the parameters without making any changes, an easy access to "help" screens that define the parameters currently on the screen, and finally a panic button that allows the user to exit the system easily at any time.

Data Linkages

The system is interfaced with a very sophisticated data management system that is beyond the scope of this paper to describe. Again, the overriding objective is to reduce data preparation time for map input. Therefore, in order to select the data for plotting, one simply enters the logical name and physical name of the file, the name of the geocode and the name of the data field. If the data is not entered in the data management system, one must define the physical characteristics of the two fields (ie. offset and length).

Geographic databases

The system has three geographic databases, namely, POINTS, WINDOW and COMMUNITY database. The POINTS database contains all of the geographic and political boundaries for the system. The WINDOW database contains definitions of commonly used windows (ie. sections of the national map which are most frequently used). The COMMUNITY database contains the co-ordinates and the names of the 1,133 standard set of communities for which DRIE maintains its COMMUNITY database.

The POINTS database (Fig. 1) is organized by geocodes at its top level and by line segments at its bottom level. As with all databases, its design is a compromise between the logical requirements of the data structure and the physical limitations of the Data Base Management System. The Geocode Master provides access to the Geocode Index, which contains an ordered list of the boundary segments that outline the specified geographic area. The Segment Master provides access to the Points file which is organized in strings of co-ordinate pairs that define one line segment. The Prov Master is a special case of the other geocode masters, in that, in addition to the Prov Index, it also provides access into the Geocode Index, thus providing access to all geographic areas within a specified province.

In operation, the system uses the data base in two distinct ways. The drawing of boundaries and shading of maps requires access to different paths through the data base. In order to draw boundaries, the system uses the appropriate Geocode Master to identify the line segments that comprise the boundary for the geocode and sets a corresponding flag in a segment mask. Once all the required geocodes have been processed, the system uses the mask to select the line segments to be plotted. The use of the segment mask prevents the segments that have multiple ownership from being plotted more than once. In order to shade the geographic areas, the system uses the ordered pointers in the Geocode Index to create polygons that can then be shaded. It should be noted that while the shading routines handle geographic areas that are comprised of multiple polygons, the system is oriented to political boundries rather than physical features.

The COMMUNITY data base contains the co-ordinates of the 1,133 communities that are contained in the COMMUNITY database. It also contains the names of selected communities for labelling purposes.

The WINDOW database contains the definition of commonly used maps. The WINDOW records contain the parameters required to define the relationship between the subject space (the selected geographic area) and the object space (the features of the geographic area to be plotted). It contains the co-ordinates for the standard

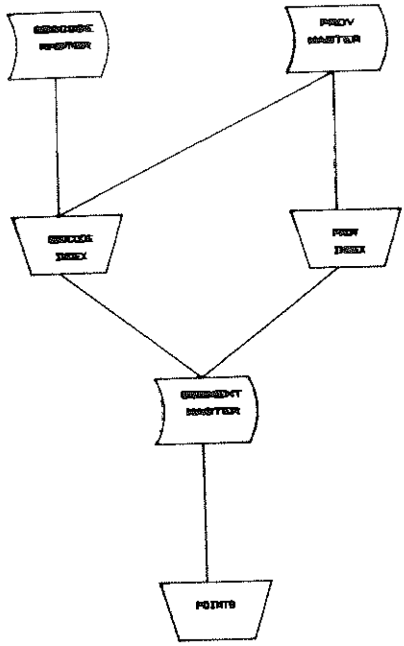


Figure 1. Points Database

location of the legend. Finally, it contains a mask defining the geopolitical areas encompassed by the window used by the system to optimise the plotting process for the selected window. Because windows are much easier to use, this feature simplifies the composition of maps. In practice, new 'standard' windows are often created for new projects.

SYSTEM CAPABILITIES

The system capabilities can be summarized as follows: the ability to define the map size and output media (page definition), the options in accessing and classifying data (data definition), and the possibilities in specifying areas to be plotted (map definition).

Page Definition

In practice, there are five standard sizes for the media that are used for map production. However, the system can produce maps of any size that will fit on the plotter. Each page that we produce can have a title placed along the bottom. All of our products have a system 'signature' on their left hand side that gives the date and map file name for the map. This 'signature' has proven to be invaluable when it comes time to reproducing copies many months after a project has been completed. A standard map will fit within the frame of a page. As well, the system is set up so that any number of maps can be placed upon the page. What sppear to be insets on our pages are in fact maps in their own right, produced with the same degree of flexibility as the apparent master maps.

Data Definition

Parameters for accessing the data have been discussed previously. In addition to those parameters the user may select from one of the ten geocoding systems currently installed in the system. They are:

1. County
2. County Ecumene
3. Labor Force Region
4. Labor Force Region Ecumene
5. Federal Electoral District
6. Federal Electoral District Ecumene
7. Community
8. Locality by Latitude and Longitude
9. Québec Administrative Region
10. Québec Administrative Region Ecumene

The user selects the number of classes of data to be used. A limit of eight has been found to be more than adequate in practice, as too many classes produce a visually confusing map which defeats the purpose of the system. The system provides seven automated classification methods as well as the option of user-defined classes. Whichever method is used to set the class intervals, the user can then select from more than 200 distinct shsding types. The system is set up so that the class definition includes the definition of the legend. The user can add titles, class description and text to the legend, or specify that certain parts of the legend not appear on the plot.

Map Definition

The final stage of definition is the selection of the area of the country to be plotted. The plotting windows can be selected by one of three methods:

1. Predefined window selection from the WINDOW database as previously described.

2. Selection of geopolitical areas of interest which the system will center within the area of window co-ordinates that it calculates.

3. Manual entry of window co-ordinates.

The system is set up so that it is not necessary to plot the same boundaries that are specified for the data classes. For example, we often specify ecumene shading but use standard boundaries on the msp. At other times we find it informative to overlay federal electoral district boundaries on county or community data.

The final specification is to select the plotting method, which with many combinations and permutations, dictates whether to have the map produced to fill the entire frame or to have the system edit out the boundaries and data between the specified geopolitical area and the map frame.

FUTURE DIRECTIONS

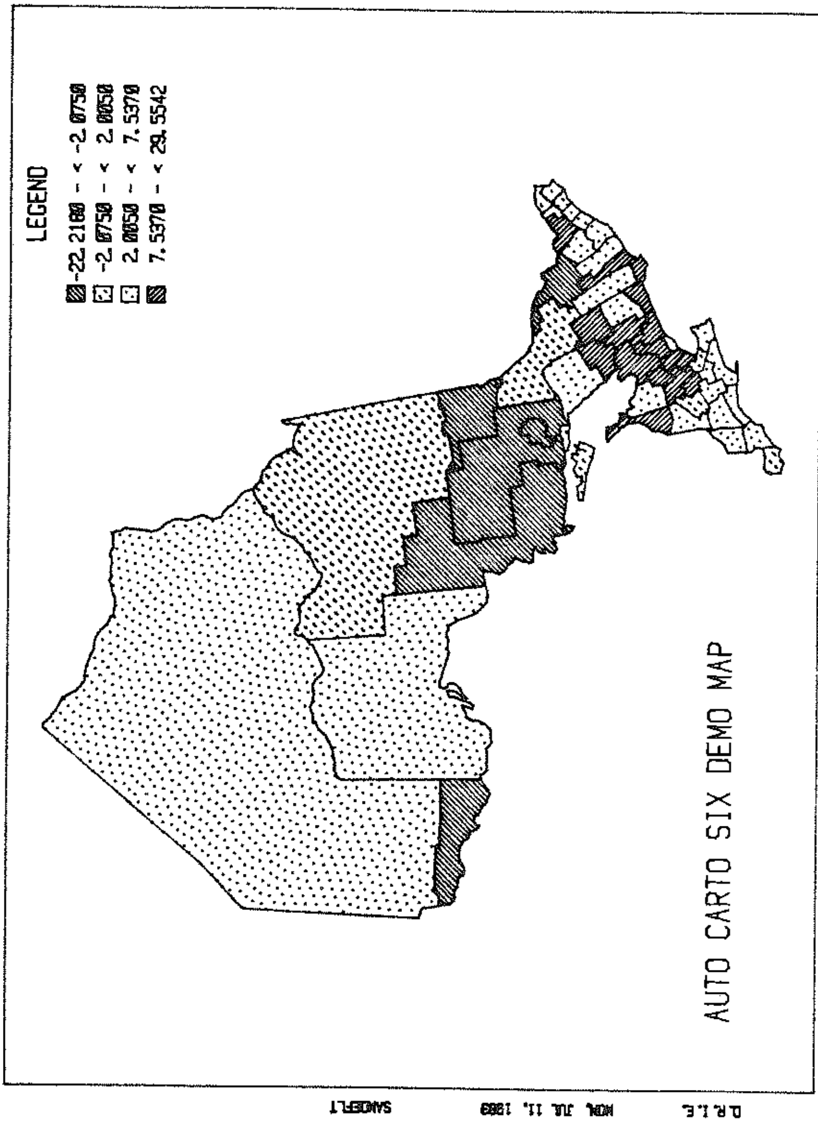
The system has been in constant development and production for the past three years. In that time we have identified several areas in which the system could be improved.

The system is now limited to production from data stored in sequential files. Although the creation of such files from our databases is automated by one of our other systems, it would be more convenient to have direct access to the databases. This change has been on our development list since version one, but has always been delayed by the requirement for the composition of more sophisticated maps.

Another enhancement would be to augment our POINTS database with more physical features. We feel that the addition of highways, lakes and rivers would give the decision makers, for whom the maps are intended, a better frame of reference than the present boundries with which many users are unfamiliar.

A major change that we would like to make to the system would be to add a set of standard graphical capabilities. The system would be integrated so that we could produce insets of line, bar and pie charts based on the class and data selection specified for the map.

Finally, and equally important, would be the acquisition of a colour CRT. In this regard, we have experimented with a black and white CRT but the results were unsatisfactory. Because of its rapid retrieval and display capability, we feel that a colour CRT would greatly reduce composition time.


LEGEND
-22.2100 - < -2.0750
-2.0750 - < 2.0050
2.0050 - < 7.5370
7.5370 - < 29.5542
AUTO CARTO SIX DEMO MAP
NET MIGRATION 1971 - 1976

Appendix 1

DESIGNING INTERACTIVE CARTOGRAPHIC SYSTEMS USING THE CONCEPTS OF REAL AND VIRTUAL MAPS

Harold Moellering
Ohio State University
Columbus, Ohio
U.S.A. 43210
Tel: (614) 422-2608

ABSTRACT

Interactive cartographic data processing systems can be designed for many different kinds of purposes. Most modern systems are organized such that they can accomplish several kinds of tasks. When one begins to lay out the design of an interactive cartographic system, one must explicitly recognize and reference different states of cartographic information in a systematic way. This can be accomplished by utilizing the concepts of real and virtual maps. The operations of such systems continually change the state of cartographic information which can be understood as transformations between various kinds of real and virtual maps. One can then combine the referencing of transformations between various states of cartographic information together with the standard modular designs of cartographic system organization. This paper is an extension beyond the earlier work of the author and will discuss how this can be accomplished. It will also show the advantages of such a strategy as an aid to conceptual understanding and as a tool to help the cartographer more effectively design such systems.

I. INTRODUCTION

During the last two decades interactive cartographic systems have developed from a laboratory curiosity to a very powerful tool for cartographic production, analysis of cartographic data and display of results. In many cases it has been recognized that these sorts of cartographic operations were changing the cartographic state of the data which was being processed. Such changes of the state of cartographic data have been clearly recognized by Moellering (1977, 1980) in the development of the concepts of real and virtual maps. When one begins to lay out the design of an interactive cartographic system, one must explicitly recognize and reference these different states of cartographic information in an explicit way. The operations of such systems continually change the state of cartographic information between various kinds of real and virtual maps.

This paper extends the earlier work of the author to show how the concept of real and virtual maps can be used as a conceptual aid in the design of interactive cartographic systems.

II. REAL AND VIRTUAL MAPS

As early work developed with cartographic data in a digital form, a number of workers in the field began to realize that although the form of information was still a cartographic, it was very different from what cartographers generally recognized as a map. One kind of new cartographic form was a CRT image which Riffe (1970) called a temporary map. Others called them ephemeral maps. Morrison (1974) recognized that with all of the new developments in cartography, it would be necessary to expand the definition of what constitutes a map. Moellering (1980) encapsulated this expansion of cartographic products and states of cartographic information into the concept of real and virtual maps. He recognized that there were two central attributes of maps which differentiated most conventional maps from these new cartographic products. These two attributes are 1) a permanent tangible reality and 2) direct viewability as a cartographic image. From these two fundamental characteristics of maps one can develop a four-fold classification of real and virtual maps. It turns out that a real (conventional) map has both of these attributes while a virtual map lacks one or both of these attributes:

> Real map. Any cartographic product which has a directly viewable cartographic image and has a permanent tangible reality (hard copy). There is no differentiation as to whether that real map was produced by mechanical, electronic or manual means.
>
> Virtual map (type I). Has a directly viewable cartographic image but only a transient reality as has a CRT map image. This is what Riffe called a temporary map. Given the direction of current scientific work, electrocognitive displays may be possible.
>
> Virtual map (type II). Has a permanent tangible reality, but cannot be directly viewed as a cartographic image. These are all hard copy media, but in all cases these products must be further processed to be made viewable. It is interesting to note that the film animation adds a temporal dimension to the cartographic information.
>
> Virtual map (type III). Has neither of the characteristics of the earlier classes, but can be converted into real map as readily as the other two classes of virtual maps. Computer based information in this form is usually very easily manipulated.

One can note that this four class system of real and virtual maps is an exhaustive system which assures that new cartographic products developed in the future will always fit into this system. Real maps are hard copy cartographic products such as sheet maps, atlases, globes, plastic 3-D relief models and the like which can be directly viewed and interpreted. All virtual maps lack one or both of these attributes of real maps. Here the virtual type I map has only a direct viewability with examples such as a CRT map image or the 2-D or 3-D cognitive images in the mind of the map reader. The virtual map type II is a hard copy product, but does not have direct viewability. Examples of this kind of map are stored holograms or Fourier transformations and laser disk data, traditional field data, gazetteer, anaglyph or a film animation. All of this class of hard copy maps must be processed in order that one may view the image contained in the product. Virtual type III map is a class of map which has neither of the two fundamental attributes and most of this kind of cartographic information is in the digital computer domain. It includes digital information stored on disk or tape, core memory, video animation as well as cognitive maps of the kind which contains relational geographic information. Notice that virtual map classes I and III contains the kinds of maps that are in the mainstream of digital cartographic processing or in the mental domain. Because these two classes of maps are very transient and lack a permanent tangible reality, the information contained in these sorts of maps is extremely manipulable, both in the digital and the graphic domains which makes them very useful in an interactive cartographic setting.

III. TRANSFORMATIONS BETWEEN REAL AND VIRTUAL MAPS

It is interesting to discover that transformations between these four classes of real and virtual maps, as shown in Figure 1, define most of the important operations in

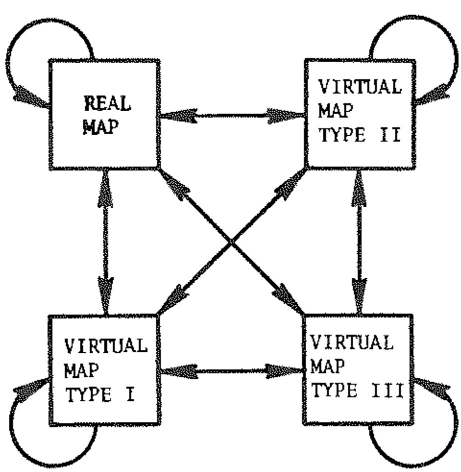


Figure 1. Real and Virtual Map Transformations

cartography. An earlier and less formal set of transformations was suggested by Tobler (1979). Most cartographic systems utilize some combination of these 16 transformations in their design and operation:

$R \to R$ Transformations of real to real maps define confine conventional cartographic operations and processes. Such tasks are either manual, mechanical or photographic.

$R \to V_1$ Traditional map reading takes place when a map reader looks at the map and creates a map image in his mind. An analog television display of a real map could also be utilized here.

$R \to V_2$ Extracting information from a map and storing it in a nongraphic hard copy form is an example of this transformation.

$R \to V_3$ Digitizing is an extremely important operation in modern cartography where information is converted from a real graphic form into virtual data and coordinates that can be utilized and stored in a very manipulable form.

$V_1 \to R$ Creating hard copy maps from graphic CRT terminals is an extremely important aspect of this transformation. It could also include creating hard copy representations of cognitive images stored in the mind.

$V_1 \to V_1$ Here one can manipulate a graphic CRT image in a number of ways, or one could read such a cartographic image and create an image in the mind.

$V_1 \to V_2$ When one converts a cartographic CRT image into a hard copy form such as a hologram, Fourier transform or film animation this transformation is used. A number of early cartographic film animations were photographed from V_1 CRT images.

$V_1 \to V_3$ Modifying a digital cartographic data base from a graphic CRT image with a device such as a cursor utilizes this transformation.

$V_2 \to R$ Using traditional field data to create a real map or any time stored hard copy data is used to create a hard copy map uses this transformation.

$V_2 \to V_1$ This transformation is used when laser disk data or perhaps field data is directly displayed in a graphic form on a CRT.

$V_2 \to V_2$ Whenever nongraphic hard copy cartographic information is converted into a different nongraphic hard copy form this transformation is utilized.

$V_2 \to V_3$ Entering coded cartographic data into a data base is an example of this transformation.

$V_3 \to R$ This transformation represents the extremely important operation of digital plotting or other form of hard copy creation such as model carving.

$V_3 \to V_1$ This transformation includes the creation of cartographic CRT displays from a digital data base which is a key operation in any interactive cartographic system.

$V_3 \to V_2$ Creating a tabular hard copy output from a cartographic data base utilizes this transformation.

$V_3 \to V_3$ This transformation defines many mathematical operations in cartography such as interchanging data between two cartographic data bases. Many of Tobler's mathematical transformations (1961) fall into this operation.

This above set of transformations defines all important cartographic transformations and hence can be used as a tool to assist in the design of interactive cartographic systems.

Surface and Deep Cartographic Structure

It is only relatively recently that cartographers have come to realize that when one desires to work with cartographic information in the digital domain that the graphic representation is only one fundamental aspect of the map information. Many and other kinds of relationships have been recognized which are not necessarily graphic. Nyerges (1980) has recognized this problem explicitly by extending the concept of surface and deep structure from structural linguistics into the cartographic domain. Surface structure of cartographic information is the graphic representation of the terminal elements which appear in the form of a map. However, there are many other explicit and implicit relationships between cartographic objects contained in the data base that are not graphic. Deep structure relates to these relationships between cartographic objects that are not necessarily graphic. These relationships can be phenomenological, spatial, in some cases nonterminal graphic elements, or a combination of these. Such relationships are not evident from the surface structure e.g. the graphic representation of the map.

Figure 2 shows a straightforward example of surface and deep structure. The lower part of the figure shows the surface structure which here is the graphic realization of the cartographic information. The upper part of the figure shows the deep structure of the cartographic information in the form of a data base schema. Many things which are seemingly not related to each other in the surface structure are in fact related as shown in the deep structure. Here logical modules of links and nodes are related to each other in a structural way which includes the intermediate points which provides drafting accuracy as well and a module of attributes of the links. Point objects which are seemingly not related in the surface structure are seen to be related via the point object module. Although this is a very simple example, it clearly illustrates the concepts of the deep and surface structure of cartographic information.

III. DESIGN OF INTERACTIVE CARTOGRAPHIC SYSTEMS

The task of interactive cartographic systems design is a challenge to all who have been involved with it and a ripe topic for conceptual discussion. It has been realized for a number of years that the man-machine interaction in a cartographic setting is a much richer information channel than conventional cartographic approaches (Moellering, 1977). More recent conceptual work on cartographic systems design continues to develop and expand the conceptual

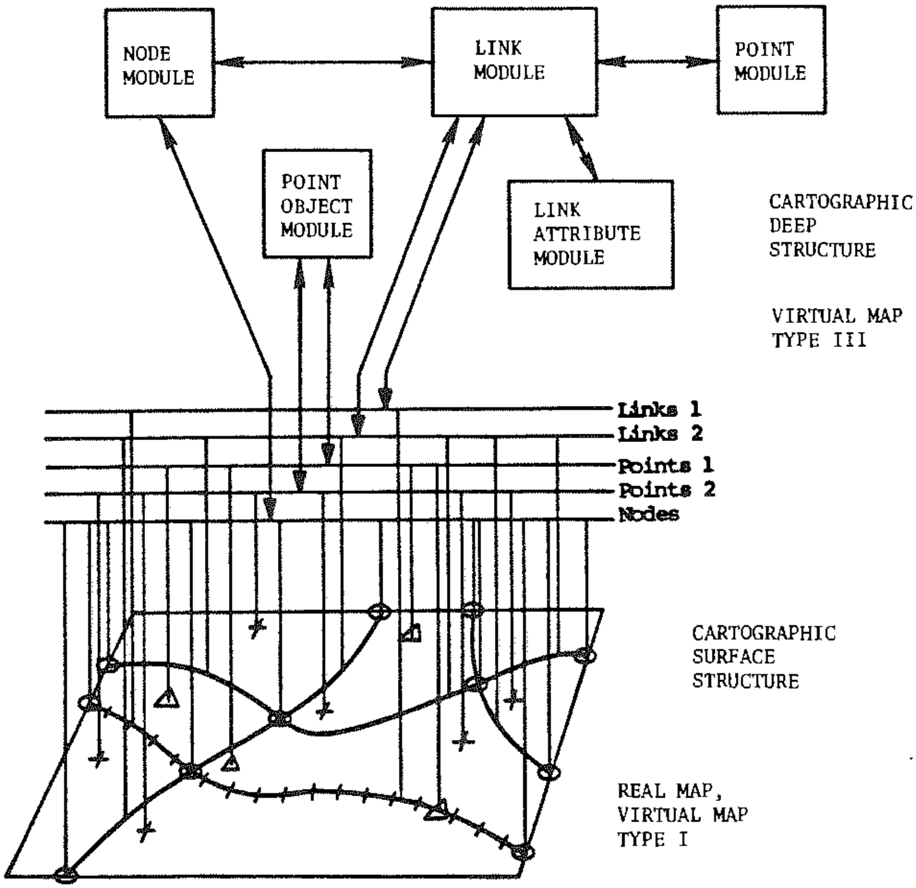


Figure 2. Representation of Deep and Surface Cartographic Structure

flexibility of such systems as evidenced by the work by Petrie (1981), Anderson and Shapiro (1979), Fischler et. al. (1979) and Pfaff and Maderlechner (1982). In any case there are five basic phases in the development of such a system: 1) specifying the system goals, 2) specifying system needs and feasibility, 3) designing the system, 4) implementing the system, and 5) system testing, verification and documentation. The early stages of this process can be developed along the lines that are similar for geographical information systems as discussed by Tomlinson, Marble and Calkins (1976). However, in this paper particular attention will be directed towards the use of the concepts of real and virtual maps in the third step of system design and assume that earlier steps in the process have been properly followed.

The use of the concept of real and virtual maps is very advantageous in the design stage of interactive cartographic systems building because it can be used to make explicit what was implicit heretofore. Now important cartographic information transformations such as $t(R \rightarrow V_3)$, $t(V_3 \rightarrow V_1)$, $t(V_1 \rightarrow V_3)$, $t(V_1 \rightarrow R)$ and $t(V_3 \rightarrow R)$ can be recognized explicitly as a transformation process in a wide variety of logical system design diagrams. It also turns out that these transformations can be shown at several levels of specificity from very general system diagrams down to very detailed logical diagrams of interactive software. They can be used to illustrate system logic, hardware configurations and software organization. In short, one can explicitly reference these transformations at any scale of system specificity or in any system environment. The adaptability and flexibility of these transformations greatly enhances the appeal of them because they are not limited by specificity or environment.

The following examples discussed here have real and virtual map transformations that are used in interactive cartographic systems design. Each of the three examples uses the transformations at a different level of specificity and in a different environment. Although only three examples cannot be considered exhaustive, they do show the flexibility and adaptability of these transformations.

Figure 3 shows a general diagram of an interactive cartographic system. It shows the entry of raw data into the system via digitizing $t(R \rightarrow V_3)$ and the tabular input of data $t(V_2 \rightarrow V_3)$. Inside the system a host of transformations can be invoked to analyze, manipulate and display the data such as $t(V_3 \rightarrow V_1)$, $t(V_1 \rightarrow V_1)$, $t(V_1 \rightarrow V_3)$, $t(V_2 \rightarrow V_3)$ and $t(V_3 \rightarrow V_3)$. Not all possible combinations are shown in this figure. After manipulation, analysis and display of the data, several kinds of cartographic products can be generated by the system as is appropriate to the problem at hand. They include generating real maps $t(V_3 \rightarrow R)$ output cartographic data structures $t(V_3 \rightarrow V_3)$ or some sort of tabular printout $t(V_3 \rightarrow V_2)$. Further details concerning the capability and operation of the system however have not been specified at this level and to do so would require more detailed diagrams.

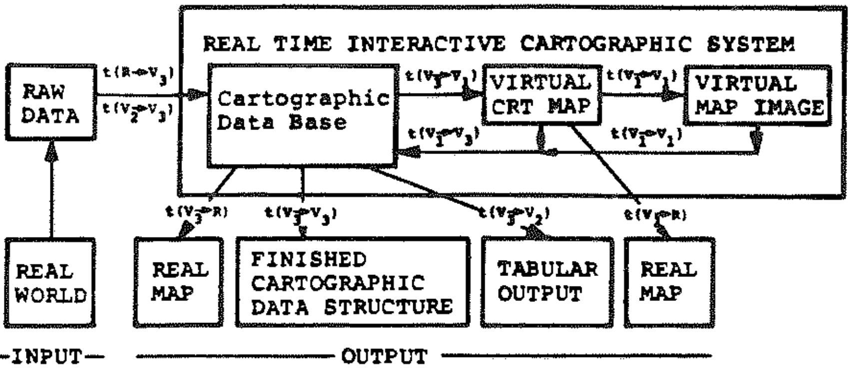


Figure 3. General Logic of Cartographic System Showing Real and Virtual Map Transformations

Figure 4 shows the hardware configuration of a single work station for a cartographic system. As depicted here it could be used for either digitizing, editing or plot verification. In some larger systems these functions might be separated into more specialized kinds of work stations. The central piece of equipment is the CRT display $t(V_3 \rightarrow V_1)$ interfaced to a host computer system with attached graphic cursor, $t(V_1 \rightarrow V_3)$, keyboard $t(V_2 \rightarrow V_3)$, $t(V_2 \rightarrow V_1)$ and function buttons, $t(V_2 \rightarrow V_3)$, $t(V_2 \rightarrow V_1)$ because they could send signals to both the screen and to the system. Several typical peripheral devices are shown in the form of a graphic tablet $t(R \rightarrow V_3)$, $t(R \rightarrow V_1)$, hard unit $t(V_1 \rightarrow R)$ and plotter $t(V_3 \rightarrow R)$. Each transformation shown illustrates the basic way in which each piece of hardware processes the basic cartographic information which passes through it. Hardware configuration design can be very useful when one is trying to assemble the proper hardware units to perform a stated combination of transformations.

Figure 5 shows a logical flow diagram of a computer program. Because interactive cartographic computer programs tend to have a large amount of man-machine communication, it is important to explicitly show these transformations as an aid to efficiently design the logic of the program. This example shows the most detailed use of these transformations and because of that the program chosen for this example is rather small. As a matter of fact, it is used to train beginners in the use of a graphic CRt terminal. The first section begins with the generating of data, the

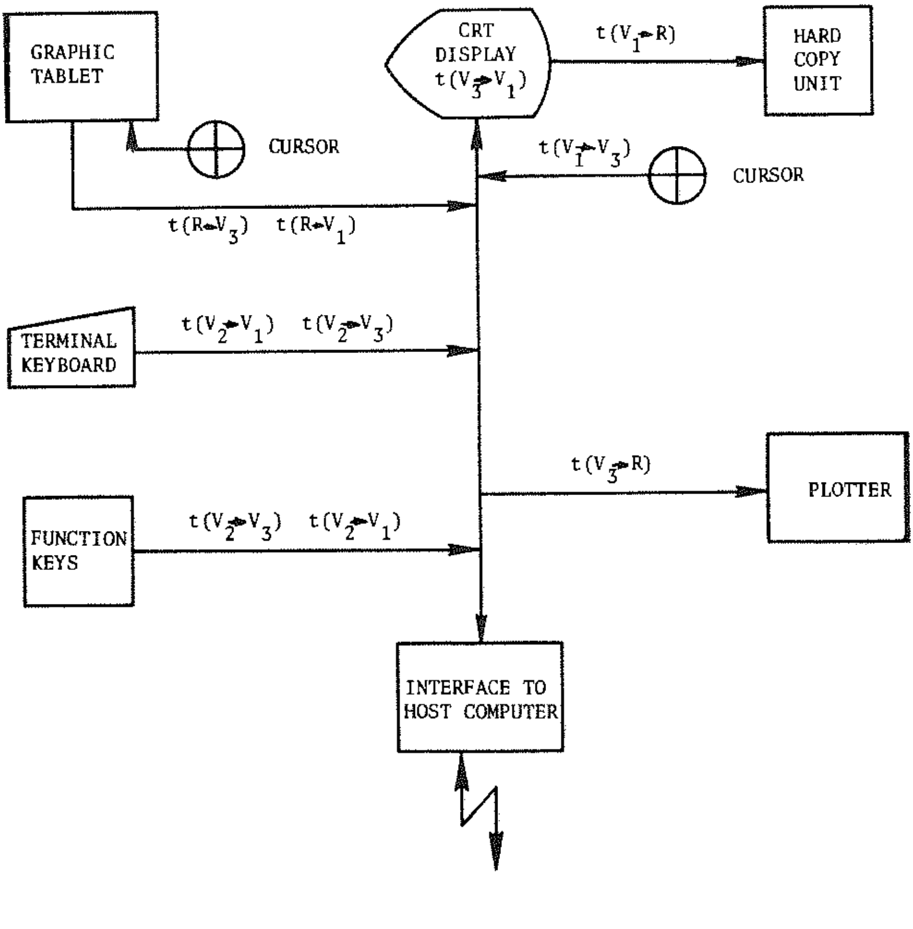


Figure 4. Hardware Configuration of Cartographic Work Station Using Real and Virtual Map Transformations

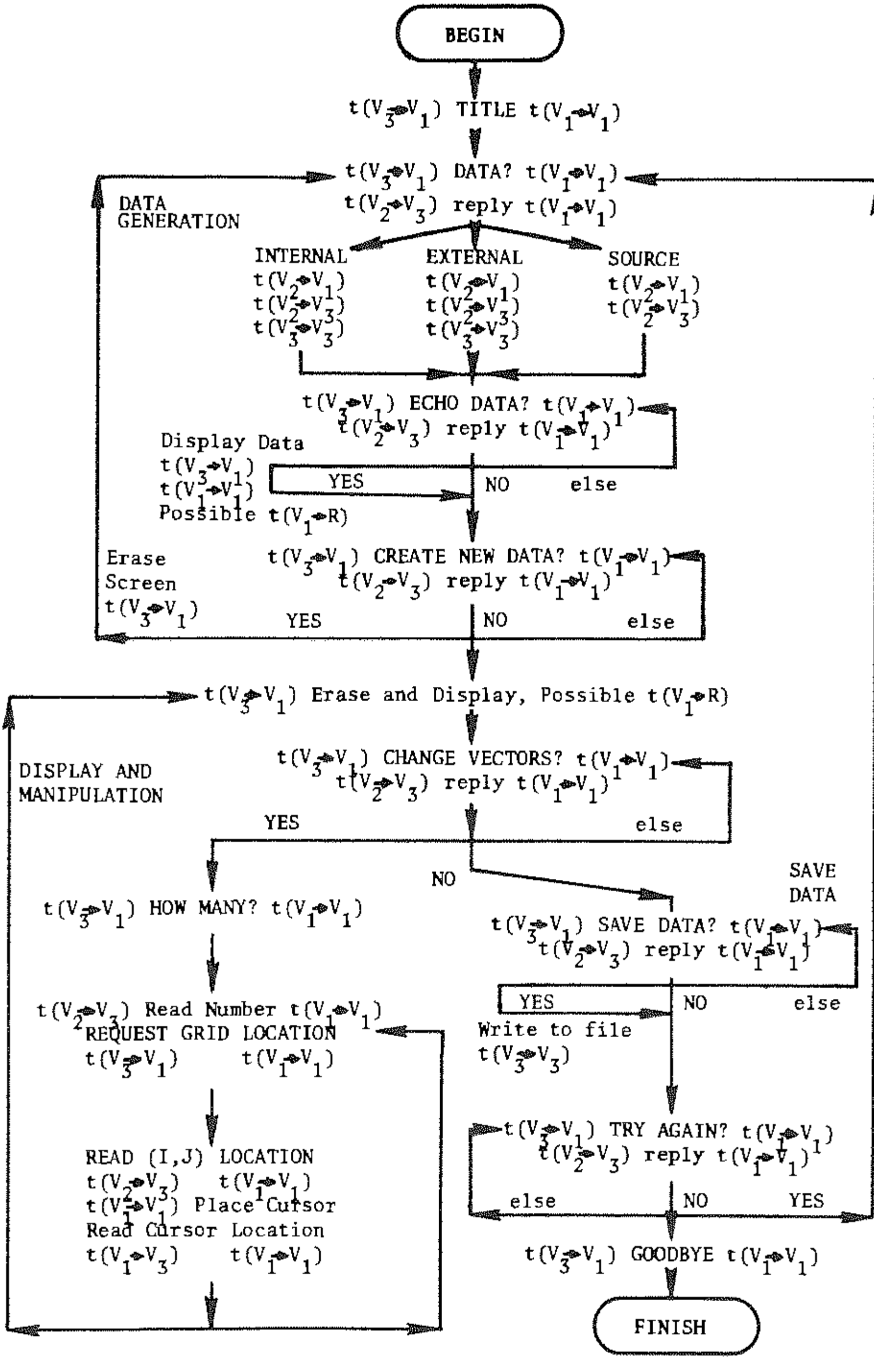


Figure 5. Interactive Cartographic System Software Logic Using Real and Virtual Map Transformations

second with the display and manipulation of the data, and the third with saving the data and terminating the program. Because there are many decisions to be made by the user of the program, a rather high level of man-machine communication is included which in turn requires that many cartographic transformations also be included. A typical command sequence is as follows: the system prompts the user for a command, $t(V_3 \rightarrow V_1)$, the user reads the message, $t(V_1 \rightarrow V_1)$, the user responds by either typing a command from a keyboard which appears on the screen $t(V_2 \rightarrow V_1)$ and is sent back to the system $t(V_2 \rightarrow V_3)$ or by pushing a function button which would either affect the screen $t(V_2 \rightarrow V_1)$ or send a signal to the host system $t(V_2 \rightarrow V_3)$, or the user could activate the cursor $t(V_1 \rightarrow V_3)$ and return the information to the system $t(V_2 \rightarrow V_3)$. Exactly which alternative of the command sequence is followed depends on the specifics of the operation to be performed. As one can see from this illustration, the program contains many of the kinds of command sequences because the user is making decisions about what the program is supposed to be doing and those decisions must be communicated to the host system and to the program software. It should be noted that this approach can be used on both individual software modules as well as entire programs. As one proceeds through the program, each of the transformations must be executed in the specified sequence or the man-machine communication could be disrupted. Therefore it is important that the software designer who works out the communication details of the program apply these transformations in the exact sequence required for efficient communication. To do otherwise could result in inefficient or perhaps unworkable cartographic programs.

IV. CONCLUSIONS.

It can be seen from the outset that the development of the concept of real and virtual maps rather neatly expands the definition of the map to include the most modern forms of maps. It can also be seen that transformations between various states of real and virtual maps specify all important cartographic operations. Further it can be seen that these transformations are a very useful aid in designing cartographic systems, especially interactive systems. The transformations can be used at a wide number of levels of specificity as well as to help define general systems logic, hardware configurations and software organization. Together with other design tools, the explicit use of these cartographic transformations should help the cartographer to design more efficient and powerful systems in the future.

V. BIBLIOGRAPHY

Anderson, Robert H., and Norman Z. Shapiro, 1979, Design Considerations for Computer-Based Interactive Map Display Systems, Rand Report R-2382-ARPA.

Fishler, Martin A, et.al., 1979, Interactive Aids for Cartography and Photo Interpretation, SRI International Report ARPA 2894-5.

Moellering, Harold, 1977, "Interactive Cartographic Design", Proc. ACSM Spring 1977, pp. 516-530.

________________, 1980, "Strategies of Real-time Cartography", Cartographic Journal, 17(1), pp. 12-15.

Morrison, Joel, 1974, "The Changing Philosophical-Technical Aspects of Cartography", American Cartographer, 1(1) pp. 5-14.

Nyerges, Timothy, 1980, Modeling the Structure of Cartographic Information for Query Processing, unpublished Ph.D. dissertation, Dept. of Geography, Ohio State University, 203 pp.

Petrie, G., 1981, "Hardware Aspects of Digital Mapping", Photogrammetric Engineering and Remote Sensing, 47(3), pp. 307-320.

Pfaff, Gunther, and Gerd Maderlechner, 1982, "Tools for Configuring Interactive Picture Processing Systems", IEEE Computer Graphics and Applications, 2(5), pp. 35-49.

Riffe, P. 1970, "Conventional Map, Temporary Map, or Nonmap?", International Yearbook of Cartography, 10, pp. 95-103.

Tobler, Waldo, 1961, Map Transformations of Geographic Space, unpublished Ph.D. dissertation, Dept. of Geography, University of Washington, Seattle WA.

______________, 1979, "A Transformational View of Cartography", American Cartographer, 6(2) pp. 101-106.

Tomlinson, R., H. Calkins and D. Marble, 1976, Computer Handling of Geographical Data, Geneva: UNESCO Press.

DESIGN OF A TELIDON BASED IMAGE ANALYSIS SYSTEM (TELIAS)

Glenn Mitchell Shirtliffe
Cartographic Research Unit
Department of Geography
Carleton University
Ottawa, Canada. K1S-5B6

ABSTRACT

The merits of digital image analysis systems in remote sensing have been well documented in the literature. The major factor that has served to limit the diffusion of digital image analysis systems throughout the research community has been that of cost.

The objective of this project was to create a digital image analysis system that would provide reasonable results at the least possible cost, using equipment already available in most computing/research facilities.

The TELIDON Image Analysis System (TELIAS) is a CP/M based system that presents digital imagery on TELIDON display devices. The interaction model is one between operator and a TELIDON display terminal, using a microcomputer as the interface. TELIDON, specifically North American Presentation Level Protocal Syntax (NAPLPS), was chosen as the base for graphic presentation primarily because of cost considerations.

INTRODUCTION

The increased amounts of digital imagery made available in the 1970s, especially from LANDSAT, provided a new source for, and a different scale of remotely sensed imagery. Preceeding the availability of digital data for image analysis, the interpretation of images of the earth's surface was possible only through analogue techniques. The two major technological prerequisites that must be fulfilled before digital image analysis can be realized are:

1. The advent of machines to collect and convert analogue data to digital format, to process the digital data, and to convert the data back to a useable analogue output.

2. The advent of machines that perform the above tasks at a price affordable to the researcher.

For major research institutions and large users, digital image processing has long been realized. Smaller users are still waiting for technology to fulfill the second prerequisite.

Early Image Analysis Systems (IAS) were implemented on dedicated mainframe systems, often costing in excess of one million dollars. Technology has progressed to the point that powerful mini-computer based systems are now available for only a few hundred thousand dollars. Unfortunately, for a majority of professional and educational institutions, limited budgets still mitigate against the acquisition of these new mini-computer based systems.

MICROPROCESSORS AND IMAGE ANALYSIS

The merits of digital analysis systems in remote sensing have been well documented in the literature. The primary factor limiting the diffusion and use of these systems has been that of cost. The development of microprocessors and their use in microcomputer systems has led to attempts to overcome the cost limitation by development of microcomputer-based image analysis systems. The last few years have seen a very rapid growth in the numbers of image analysis systems available for microcomputers.

There are three types of image analysis systems available at present, each with differing capabilities, design complexities, and greatly differing price tags. One factor however, that is common to all of these micro-based systems, is the trade-off of processing power to cost. Simple digital image analysis (DIA) functions that take a matter of minutes to execute on mainframe installations, may take several hours to process on a micro-computer. Obviously, a careful division of labour between mainframe and micro is called for.

The most complex and costly micro-based image analysis systems are dedicated 8 or 16 bit systems with configurations and hardware specially designed for image processing and display. They range in price from $15,000 to $50,000 and have more complex software enabling the user to perform many of the functions available on mainframe systems. The second type of system is in the $5,000 to $15,000 range. Such systems tend to be powerful 8 bit machines with dedicated graphics memory and some sophisticated image processing functions. The third type of system is priced under $5,000. These tend to be software packages for the personal computer, or system modifications to the personal computer. Such systems have only limited processing capabilities, and very limited display capabilities. The systems reviewed below are examples of systems in the latter two categories.

IMPAC/RIPS. The Image Analysis Package for Microcomputers (IMPAC) developed by Egbert Scientific Software, and the Remote Image Processing System (RIPS) developed by the EROS Data Center are examples of systems in the middle price range. The IMPAC system is designed on a Cromenco Z-80 microcomputer with 64k RAM, and 96k image memory. It has 256x240x12 bit resolution and displays on a Panasonic RGB monitor. Hardcopy is provided by a Data Systems Prism 80 dot matrix colour printer. IMPAC has two 5 1/4 inch disk drives, providing 630k of storage. The software is implemented in CBASIC on a CP/M operating system. Utilities include image algebra, and density slicing algorithms as well as the following classification algorithms: parallelepiped, maximum likelihood, and unsupervised clustering.

RIPS is similar to IMPAC in its hardware configuration, the major difference being it uses two 8" double-sided double-density disks for 2.4 Mbytes of storage. RIPS is implemented in FORTRAN, and includes routines for contrast enhancement, band ratioing, athmospheric haze removal, density slicing, and spatial filtering. Various supervised and unsupervised classification algorithms are provided.

The primary market of these two systems is the educational and research community, although they have had implementations in resource management.

APPLEPIPS. The Apple Personal Image Processing System (APPLEPIPS) software package is an example of a low price range system. APPLEPIPS was developed by the Telesys Group, as a software package for the APPLE II or II+ microcomputer with 48k RAM, and two disk drives with 140k of memory. The display resolution of APPLEPIPS is 40x40 with sixteen colours, or 140x96 with six colours. The system has been implemented in BASIC and 6502 Assembly language. Utilities included in the system are single band display, single or double band image algebra, and bandwise pixel normalization. The classification utilities include a supervised classification algorithm which is based on a parallelepiped classification. APPLEPIPS is a low-cost system that will be most widely used in educational institutions that already have an APPLE computer. It is an effective tool in teaching the basics of interactive image processing. It is well documented and user friendly to those with very limited image analysis experience. This system and others in this price range, are probably too limited for applications outside of education since their display, processing, and output capabilities are greatly limited in resolution and manipulative ability.

SUMMARY. To the average educational institution or small firm, there is still a void in the availability of systems at the low end of the price range that are capable of providing reasonable minimum image analysis requirements. Such minimum requirements may be identified as follows: the system should have the capability to display images that correspond directly to existing topographic maps at either the 1:25000 or 1:50000 scale (implying greater than 200x200 pixel resolution if one seeks to display LANDSAT imagery); the ability to manipulate, process, and classify imagery using single and multiband techniques; the ability to annotate and superimpose geographic information on the image; a "zoom" capability; and hardcopy output.

It was with these system requirements in mind, and with the added constraint of using only equipment currently available within the Cartographic Research Unit at Carleton University, that the TELIAS PROJECT was born.

TELIAS

The TELIDON Image Analysis System (TELIAS) was implemented on a Dy4 Systems' ORION V, Z-80 based, microcomputer. The interaction model is one between operator and NORPAK Mark IV TELIDON (North American Presentation Level Protocal Syntax - NAPLPS compatible) decoder using the microcomputer as the interface. NAPLPS technology was chosen for a variety of reasons:

1. It employs a standard RS-232C interface between processor and display device, adding to the transferability of the image outputs.

2. The NAPLPS protocol is becoming more widely accepted as a protocol for information transfer, and it includes other graphic functions that can easily be implemented to provide text and graphic annotation.

3. The display resolution of NAPLPS meets minimum system requirements defined above.

4. The cost of the NORPAK Mark IV TELIDON decoder was far less than other colour graphic display terminals, and it can be attached to either RGB monitors, or via the RF output, to normal colour televisions.

TELIAS was designed to be as machine independent as possible. It is hosted by the CP/M operating system, using the LST: port for image output. The system is thus compatible with any 64k CP/M based micro with two double-density drives and RS-232C ports. Monochrome hardcopy is obtained from a Digital DECWRITER III. Up to 64 greytones may be obtained through overprinting, and the pixel size of the printer when set to 16h x 12v characters per inch provides accurate scaling horizontally and vertically. Figure 1 is an example of hardcopy output of LANDSAT Band 7 data from a 226x226 subscene of the White Bear Lake East region of Minnesota.

System software was designed to maximize efficiency yet remain user friendly and user apparent. For this reason, C - a 'not-very-high' level programming language - was chosen for much of the code. A UNIX-like 'shell' for directed input and output pipes adds to the efficiency of data transfer between modules of the system. Utility routines and display functions were written in Z-80 ASSEMBLY language, and were designed to be user transparent.

It is expected that many institutional and small business users will already possess much of the hardware required for TELIAS. If the system was, however, to be acquired in its entirety, the probable hardware costs would be as follows:

64k Micro w/CP/M 2SSDD Drives	$4,000
TELIDON Decoder	$1,300
RGB Monitor	$ 700
Printer	$1,500
TOTAL	$7,500

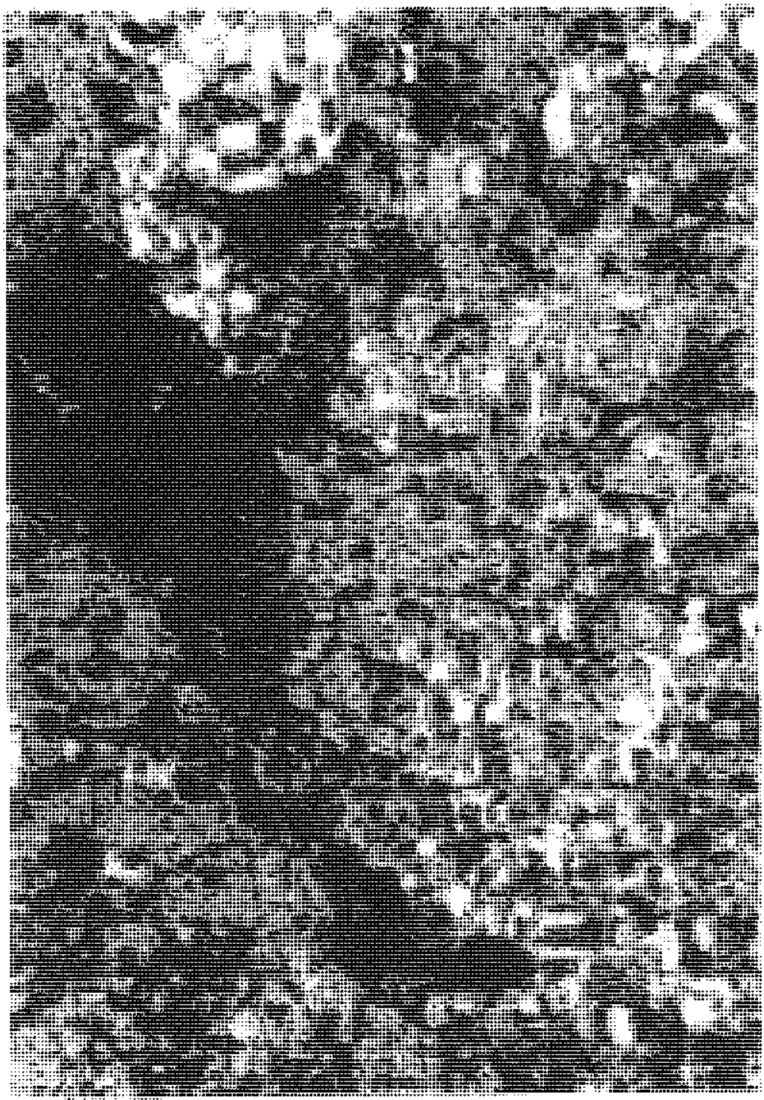


FIGURE 1. HARDCOPY OUTPUT OF LANDSAT BAND 7 DATA.

DESIGN PRINCIPLES FOR TELIAS. The five processes involved in digital image analysis are data input, pre-processing, classification, post-processing, and image output. TELIAS has been designed to facilitate all stages, however the complexity accorded to each stage varies.

The input and pre-processing stages are implemented primarily to accept digitally corrected data on CP/M compatible floppy disks. No facility is provided at present, to geographically correct, calibrate, or resample raw datasets.

In the rational division of labor between mainframe and micro, these stages are best handled by the mainframes. The availability of digital image datasets on floppy disks is increasing. The EROS Data Center now has made available inexpensive 8" CP/M format floppy disks containing LANDSAT data corresponding to individual USGS 1:24000 topographic quadrangles. In Canada, the CCRS has no firm plans as yet regarding the release of digital data on floppy disks.

There is a utility in TELIAS that enables the capture of raw image data with only slight modification to the printer, and with the addition of an A/D converter to the system. A phototransistor is attached to the printhead of the DECWRITER, and the system repeatedly spaces the printhead across an image and scrolls downwards, sampling the voltages from the transistor at each position. The density of the scan is 16h x 12v pixels per inch. This utility provides a very crude but cost effective method of digital data collection.

The classification process has received a fairly complex implementation on TELIAS. At present, the following classification algorithms have been implemented: contrast stretching (linear and non-linear); band ratioing; algebraic transformations (single and multiband); and contrast enhancement. Figure 2 is a direct-from-screen photograph of a contrast stretch of the LANDSAT Band 5 image of White Bear Lake, Minnesota.

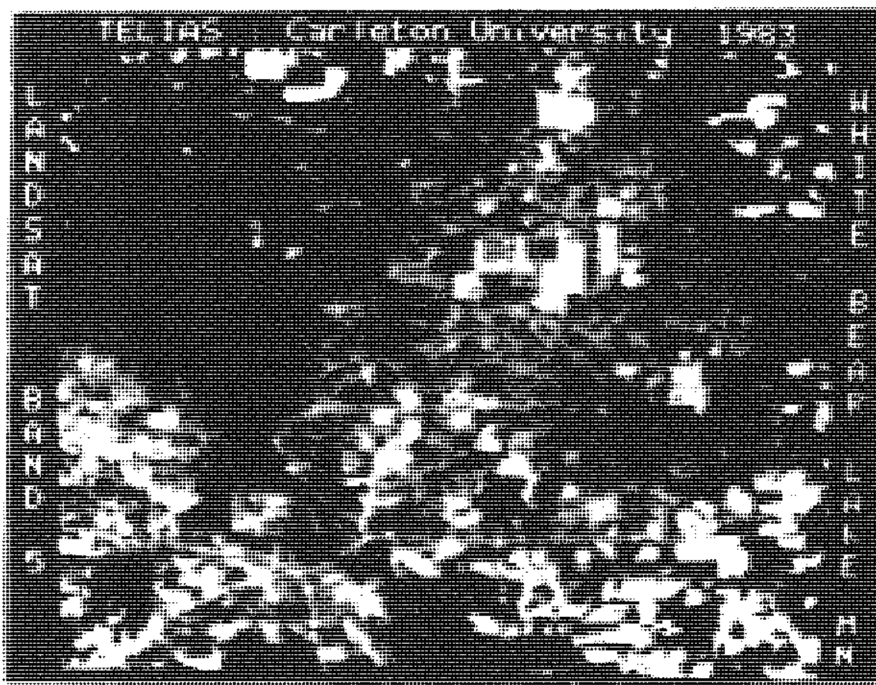


FIGURE 2. DIRECT FROM SCREEN PHOTOGRAPH OF TELIAS DISPLAY.

Other algorithms currently being developed for TELIAS are: spatial filtering; and various supervised and unsupervised classifications including maximum likelihood, minimum distance, parallelepiped, and limited-sample singular decomposition algorithms.

In the post-processing stage, the classified data is assigned intensity or hue values according to class. A software "zoom" may be specified in the video display by redefining the logical pel size used in the NAPLPS protocol. The output stage involves either the interactive display or the dumping of image files to the selected output device.

The major limitation of the TELIDON decoder in the presentation of image data is in its uniplanar display matrix. An explicit colour map defines 16 colour values, and each pixel may only assume one of these values; therefore false colour composites or band composites must be handled in software as opposed to hardware. One cannot simply assign one band to each colour gun to produce a composite. This trait is, however, not uncommon in other low cost systems.

TELIAS also has the ability to annotate and to superimpose geographic information on the image using the graphic functions of the NAPLPS protocol. There is direct compatability between TELIAS and existing GIS software written at Carleton University. MIGS, the Microcomputer Interactive Graphics System, provides a mapping capability that can be used by TELIAS to access geographic information.

<u>SYSTEM ANALYSIS</u>. The TELIAS system has met its design objectives. It is a medium resolution, low-cost, quick-look system that facilitates rudimentary image processing. It provides both video and hardcopy output, and it has the capability to collect data as well as process existing data. TELIAS can access other GIS databases and overlay geographic information.

The complete system may be acquired for under $8,000, although most users would have only to purchase the NORPAK decoder, A/D converter, phototransistor, and possibly a colour monitor - an outlay of approximately $2,500.

TELIAS is slow when compared to its mini and mainframe counterparts. Depending on the complexity of the classification it may take from 5 to 10 minutes to display an image. This length of time does compare favorably to other systems in the $5,000 to $15,000 range. The moderate resolution of the display may be improved by using a NORPAK 512x512 pixel display board, but at greater cost.

The advantage of NAPLPS technology is in its machine independence. The protocol concept of a unit screen is completely independent of the display resolution of the decoder, thus software is easily transferable to higher resolution display devices.

The greatest limitation that TELIAS and all other microcomputer based systems have, is that of power. The more advanced data classification algorithms are impractical and beyond the scope of present 8 bit microcomputers. With 16 bit micros and 32 bit minis becoming less expensive, it may not be long before this limitation is overcome.

CONCLUSION

A TELIDON - microcomputer based digital image analysis system is feasible. TELIAS is a first attempt at such a system, and the results of the project encourage further development.

Besides its application as a stand-alone image analysis system, TELIAS has also also proven that digital imagery may be effectively transferred and presented by TELIDON. This in itself may provide digital image capability to the lowest budget user by bypassing the need for a host microcomputer. A videotex database of digital imagery could easily be established on a host mainframe or minicomputer. The user would only need a decoder, colour television/monitor, and a modem to be able to access the database (approximate cost $2,500). The user could conceivably be accorded limited processing capabilities on such a database. The economics of such a videotex service may be unrealistic in the public realm, however it may provide a useful utility in large institutions or businesses where image analysts need a quick look, low resourse demand capability for planning analysis sessions on the mainframe system. Such applications warrant further investigation.

Digital image processing is rapidly becoming an affordable and useable proposition for many small scale users. TELIAS has attempted to draw together some of the more recent data processing innovations to hasten this evolution. The future of this system appears to be quite promising.

ACKNOWLEDGEMENTS

Major funding and support for the TELIAS Project was provided by the Department of Energy, Mines, and Resources Canada. The TELIAS Project was researched under the direct supervision of Dr. D.R.F. Taylor, Cartographic Research Unit, Carleton University.

REFERENCES

Bown, H.G., and C.D. O'Brien. Picture Description Instructions for the Telidon Videotex System. CRC Technical Note 699-E, Communications Research Centre, Dept. of Communications, Canada. November 1979.

Canadian Standards Association. Preliminary Standard T500-1982: Videotex/Teletex Presentation Level Protocol Syntax (NAPLPS). Standards Division, Canadian Standards Association. August 1982.

Chang, E.J.H. _Design Considerations for the Telidon-Apple Picture Creation System_. Dept. of Computer Science, University of Victoria, Canada. February 1982.

Danielson, R. Selecting An Image Analysis Minicomputer System. The 1981 Conference on Remote Sensing Education, May 18-22, 1981, Session 5c, Purdue University.

DIPIX Systems Limited. Product Specification LCT-11 Image Analysis System, Ottawa, 1982.

Godfrey, W.D.,and E.J.H. Chang. _The Telidon Book_. Victoria: Press Porcepic. 1981.

Goodenough, D.G. The Image Analysis System (CIAS) at the Canada Centre for Remote Sensing. _Canadian Journal of Remote Sensing_, Vol. 5, No. 1, May 1979, pg. 3-17.

Jensen, J.R., F.A. Ennerson, and E.J. Hajic. An Interactive Image Processing System for Remote Sensing Education. _Photogrammetric Engineering and Remote Sensing_, Vol. 15, No. 11, November 1979, pg. 1519-1527.

Peet, F.G., and J.M. Wightman. A Poor Man's Digital Image Interpretation System. _Canadian Journal of Remote Sensing_, Vol. 4, No. 1, April 1978, pg. 29-31.

Salter, L., G. Maincent, and D. Sarrat. SEP Image Analysis System: VIPS 16/65. Societe Europeene De Propulsion, Vernon France, 1981.

Spectral Data Corporation. Remote Image Processing (RIPS) Specifications. Hauppauge, New York, September 30, 1982.

Spectral Data Corporation. The SPEC-DAT Interactive Digital Image Processing System Software Description. Version 2.0, Hauppauge, New York, 1982.

Welch, R.A., T.R. Jordan, and E.L. Usery. Microcomputers in the Mapping Sciences. _Computer Graphics_, Vol. 6, No. 2, February, 1983, pg. 33-42.

INSTITUTIONAL MAPPING EFFORTS - STATUS AND PROSPECTIVE

A Report on the Benchmark Testing of a Prototype Enhanced Stereoplotter Workstation 75
Rob John

Status of Auto Carto at the Defense Mapping Agency 85
F.C. Green

Une base de données relief a l'IGN France 90
D. Piquet-Pellorce et A. Bernard

Fabrication de bases de données à référence spatiale à partir de données numériques de la carte de base du Québec à l'échelle 1:1000 99
Gervais Pellerin

The Future in Terrain Elevation Data Processing at the Defence Mapping Agency 110
Robert Edelen

Cartographie assistée par ordinateur pour le recensement du Canada 117
Ross Bradley

Hydrographic Survey Requirements System (HYSUR) 127
Chung Hye Read

Computer-Assisted Cartography for Census Collection: Canadian Achievements and Challenges 135
Joel Yan and Ross Bradley

L'utilisation des systèmes graphiques interactifs en cartographie 147
Dominique Bruger

Automated Standard Nautical Chart Production 155
Larry Strewig, Joseph Ruys and Jan Schneier

A Stereo Electro-Optical Line Imager for Automated Mapping ... 165
J.R. Gibson, R.A. O'Neil, R.A. Neville, S.M. Till and W.D. McColl

Underlying Requirements and Techniques of a Geographic Data Base (Abstract) 177
Christine Kinnear and David Meixler

Kern MAPS 300 (Abstract) 178
Jonathan Hescock

Geographic Data Procesing in Forestry - Does it Pay? 179
J.A. Benson

Long-Term Plans for the Geographic and Cartographic Support to the U.S. Bureau of the Census (Abstract) 184
Frederick Broome, Timothy Trainor and Stephen Vogel

A REPORT ON THE BENCHMARK TESTING OF A PROTOTYPE ENHANCED STEREOPLOTTER WORKSTATION

Captain R. B. John
Royal Australian Survey Corps (RASVY)
Army Survey Regiment, PO Box 226, Bendigo, Australia 3550

ABSTRACT

1. A Wild B8 stereoplotter has been enhanced by the addition of interactive, raster monochrome graphics, graphic image superimposition and voice recognition and response hardware. That new configuration has undergone a two month benchmark test and development period. The results of that period were very favourable and suggest greatly improved stereoplotter throughput in comparison with current RASVY digital techniques. The enhancement costs can be amortised in less than four years. The results of preliminary studies and of the benchmark period are reported.

INTRODUCTION

2. The Australian Department of Defence (Army Office), on 31 December 1981, entered into a contract with Intergraph Corporation for the supply and installation of a computer aided mapping system. That system is known as AUTOMAP 2. It involves the second phase of the application of computer principles to the RASVY map production system.

3. A major part of the AUTOMAP 2 contract required the contractor to develop an enhanced stereoplotter workstation, to RASVY specifications (patent pending), that would overcome many of the inefficiencies of the AUTOMAP 1 configuration. In that system, Wild B8s have been fitted with encoders and software interaction is via a VDU. A pencil plot in the model space is used to gauge progress and help with line joins. Inefficiencies of that system include a lack of digital feed back leading to inability to relate the position of the stereoplotter measuring mark to that of a previously digitized point or line and a frequent requirement to look away from the analogue model to enter and verify data and commands. Together with resultant operator uncertainty and over-cautiousness, these factors lead to reduced stereoplotter throughput and to the production of "unclean" digital data which must be verified by hard copy and digitally edited to cartographic standards.

4. The Commonwealth opted to provide optical design consultancy and prototype development for the stereoplotter workstation through the Defence Research Centre, in Salisbury, (DRCS) South Australia.

PROTOTYPE CONFIGURATION

5. RASVY and DRCS produced an optical design known as Graphics Superimposition Oculars (GSOC) which was attached to a Wild B8 stereoplotter to enable a digital image to be introduced into the left optical train and be superimposed over the left picture image.

6. A 19" diagonal, monochrome, raster graphics tube was included to display the digital data as it was collected. A second 7" diagonal, monochrome, raster graphics tube was configured to "slave" from the first and provide the line and cursor spot resolution required for

graphics superimposition.

7. Interaction between the operator and the software system is provided for in a number of ways:

a. A keyboard is provided for alphanumeric entry, to be used to "boot" the workstation and to enable miscellaneous command and data entry not available through other means.

b. A menu table is provided to enable the positioning of command and feature menus within easy reach of the operator. Interaction is via a hand held floating cursor.

c. A twelve button menu pad is provided for positioning in the proximity of the stereoplotter tracing stand (to the operator's preference) for command entry.

d. A foot switch enabling three separate data entry commands is provided.

e. A voice response system is provided to augment the alphanumeric feed-back available within the graphics superimposition field of view (one line of 60 characters).

f. A voice recognition system is provided to enable command and feature type entry without requiring the operator to turn away from the stereoplotter eyepieces.

8. In order that the superimposition of the digital image would overlay the analogue image of the left aerial picture, a sophisticated mini-computer driven software and hardware system is provided to ensure that the operator can "pan" through his model and retain that digital/analogue relationship. Operations are provided to display the digital data at any scale and to enable scale matching with that as viewed through the ocular system.

9. Software has been developed to enable the centre perspective transformation of the digital data to fit the left aerial picture but was not delivered during the benchmark period. The digital and analogue views would therefore only match exactly, at the centre of the stereoplotter field of view (when parallax had been removed by way of the Z wheel) and image overlay was dependant upon the nature of the terrain.

PREPARATORY STUDIES

Work Studies

10. In order that typical operator actions on both the analogue and Automap 1 stereoplotter configurations could be ascertained, for use in ergonomic comparisons, a work study was performed on each system.

11. Averaged results are presented in Tables 1 and 2. These figures represent the breakdown of work for typical stereomodels. A "weighted" average is calculated where the weight factors have been taken from other studies (see Table 4) and thus more closely represent true model breakdowns. (Note, however, that there is no direct relationship between the results in each table due to different procedures involved - see para 3.)

Action	Cultural	Drainage	Relief	Vegetation	Average/model (1)
Plot	71	68	73	64	70
Check	8	20	15	16	15
Enhance	13	5	10	15	10
Pencil	8	7	2	5	5
Total	100	100	100	100	100

Note 1. These figures derived by weighting percentage times according to figures in Table 4.

TABLE 1 - ANALOGUE STEREOPLOTTER WORK STUDY
BREAKDOWN OF WORK FOR TYPICAL STEREOMODELS
(Percentage Time Involved)

Action	Cultural	Drainage	Relief	Vegetation	Average/model (1)
Plot	73	70	49	57	62
Keyboard	12	6	10	9	9
Quick Check	5	8	10	3	7
Check	3	9	20	0	11
Enhance	3	3	3	27	6
Pencil	4	4	8	4	5
Total	100	100	100	100	100

Note 1. These figures derived by weighting percentage times according to figures in Table 4.

TABLE 2 - AUTOMAP 1 STEREOPLOTTER WORK STUDY
BREAKDOWN OF WORK FOR TYPICAL STEREOMODELS
(Percentage Time Involved)

12. The figures from the above tables have been used to estimate the maximum savings that are possible in relation to actual stereo-plotting time. The enhancements considered and their estimated affect on model throughput are:

a. The use of menus for command and feature selection rather than keyboard entry (-1%).

b. Direct Z encoder sampling to alleviate the requirement to enter height values when contouring or spot heighting (-2%).

c. The provision of a monochrome raster graphics tube for digital feedback and to enable interactive corrections (-15%).

d. Graphics superimposition to alleviate the requirement to turn away from the stereomodel and give better feedback on the completeness and correctness of digitization (-7%).

e. Voice recognition and response to enable command and feature selection and verification, without requiring the operator to turn away from the stereomodel (-2%).

13. Table 3 tabulates the potential model plotting saving (against AUTOMAP 1 throughput) that may be realized for each of a number of alternative configurations that could be introduced in AUTOMAP 2.

Alternative Configurations	Savings as per Para 12 a	b	c	d	e	Total Saving
B8 + Graphics	1%	2%	15%	-	-	18%
B8 + Graphics + Voice	1%	2%	15%	-	2%	20%
B8 + Graphics + Superimposition	1%	2%	15%	7%	-	25%
B8 + Graphics + Voice + Superimposit- -ion	1%	2%	15%	7%	2%	27%

TABLE 3 - POTENTIAL MODEL PLOTTING SAVING PER CONFIGURATION
(Percentage Compared To Automap 1 Timings)

Timing Studies

14. In order that total throughput timing for average models produced within the AUTOMAP 1 system could be ascertained, a study was performed over twenty-four model areas. Timings included model orientation, all detail plotting, operator and supervisor check times, final verification plotting and post-plotting edit of the data to ensure that internal and external linework joined to cartographic standards. No intermediate verification plotting was included as those times overlap with model plotting.

15. Time sheets were compiled by all machine operators, system managers and editors involved in the production of those twenty four models. The time sheets were analysed and the resultant data is tabulated in Table 4.

Stages of Stereomodel Digitization	Stereo-Plotting Require-Ment	Stereo-Plotter Operator Require-Ment	Stereo-Plotter Station Require-Ment	Total Model Require-Ment
MODEL ORIENTATION	-	8	7	5
Cultural Plot	21	18	17	12
Drainage Plot	33	28	27	19
Relief Plot	35	30	29	20
Vegetation Plot	11	10	9	7
TOTAL PLOTTING TIME	100%	(86)	(82)	(58)
Model Checking		6	6	4
TOTAL OPERATOR TIME		100%	(95)	(67)
Final verification Plotting			5	3
TOTAL STATION TIME			100%	(70)
Internal/External Line Joins				30
TOTAL MODEL TIME				100%

TABLE 4 - AUTOMAP 1 MODEL PRODUCTION TIMES
REDUCED TO PERCENTAGES

Conclusions of Preliminary Studies

16. Combining the results of the two preliminary studies described above, it was estimated that enhancement of the AUTOMAP 1 stereo-

plotter configuration could provide a model throughput saving of 41% of the current requirement. Of this saving, 16% would come from a reduced plot time at the stereoplotter while 20% would come from the alleviation of the requirement for post editing of the model data. These figures assume that the plot time would be increased by not more than 10% of the total current requirement due to the introduction of "snapping" operations. Table 5 tabulates the potential model throughput savings.

Potential Areas For Savings	Average Current Requirement (From Table 4)	Estimated Potential Saving (% of Total Time)
Model Orientation	5	2
Plotting	58	16
Checking	4	Nil
Verification Plotting	3	3
Gross Error Edit	30	20
Total	100	41

TABLE 5 - POTENTIAL MODEL THROUGHPUT SAVINGS OVER THE CURRENT AUTOMAP 1 REQUIREMENT (%)

BENCHMARK TESTING

17. A model area was selected which displayed as much variety in topography as possible. That model was replotted during each week of the development period. It was recognised at the outset that replotting the same model over and over again introduces an aspect of operator familiarity that would not be present in a normal production environment. This was seen to be advantageous since the introduction of complete model familiarity reduces the digitizing task to a level at which model throughput is completely dependent upon the ability of the operator and the digitizing system.

18. Comments relating to the use of interactive graphics, image superimposition and voice recognition and response are given below from experience received with those equipments during the benchmark period:

a. Interactive Graphics (without image superimposition).

(1) The introduction of interactive graphics provides immediate verification of digitization and leads to improved operator confidence. The graphics screen provided was a raster, monochrome, 19" diagonal screen of 1024 x 1280 pixel resolution. Real time zoom and pan facilities enabled very efficient model checking over the entire model area. Software selectable line thickness and output styles (eg dots, dashes etc) allowed for a wide choice of line symbolization adding to data depiction and verification possibilities. Four programmable prompt fields at the bottom of the screen provides textual feedback to the operator concerning feature selection, digitization requirements, system status and error information.

(2) The placement of the screen was to the immediate right side of the operator at eye level beside the stereoplotter. Minimum head movement was then required for quick glances at the screen during digitizing to verify

data capture. Even so, the operator found that this continual sideways head movement was more tiresome than in the AUTOMAP 1 configuration where up and down head movement is required to glance at an analogue plot in the model space.

(3) The use of a monochrome graphics screen, rather than colour, is not believed to be detrimental at the stereoplotter due to the wide range of line symbolization available. Experience with the system will prove (or otherwise) the ability of the operator to identify feature types from symbolization (rather than in combination with colour).

b. Graphics Image Superimposition.

(1) The benchmark successfully proved the concept of graphics image superimposition in the left optical train. The digital image was able to be superimposed over the left picture image with precise alignment at the position of the stereoplotter measuring mark. Accurate rotation and scaling of that image is performed by an interactive routine requiring operator action to identify points on the photograph using both the stereoplotter measuring mark and the cursor associated with the graphics screen. Alignment of the CRT spot (which is always in the centre of the screen) with the stereoplotter measuring mark is performed by mechanical means. A rheostat was provided to change CRT brightness to suit the image quality of the photographs and to enable the superimposed view to be turned off or displayed at operator preferred brightness.

(2) To enable the digital image to fit the left picture image exactly, the digital data must undergo a perspective transformation and be dynamically panned with any movement of the stereoplotter viewing microscopes. This software was still in development during the benchmark period. Even without this precise image alignment, the superimposed view was found to be extremely useful to the operator. It provided immediate verification of digitization and allowed precise alignment of the measuring mark to any previously digitized data (since image alignment is precise at the measuring mark even without perspective display).

(3) The line symbolization was found to be of good quality although at times the thinnest lineweight was difficult to distinguish. The hardware/software system provided excellent image movement with movement of the tracing stand. Image movement (and associated flicker) was reduced to be almost unnoticeable at the centre, and noticeable but instantaneous at the extremities of the field of view.

(4) Image superimposition of dense digital data was found to be detrimental to stereo-acuity. This was particularly so for the plotting of contours where the requirement to "look ahead" to determine the path on which to move the measuring mark relies on good height perception by the operator. The existence of a digital linework pattern over the analogue image has an adverse affect on height perception (and image merging). The floating mark can still be accurately placed "on the ground" at any

particular model location and the capture of linear features of varying Z value does not seem to be affected to the same degree. To avoid this problem, the system can be programmed to display only the feature being captured, singly or in combination, with any user specified features already captured. The operator then has the facility to display only those features required for his current digitization requirement.

(5) The prompt fields displayed at the bottom of the graphics screen are also visible within the stereoplotter field of view to provide visual feedback of feature selection, digitizing mode, system status etc.

c. Voice Response.

(1) A voice response system was provided that enabled data displayed in any of the prompt fields to be audibly repeated via a headphone set. The repeated words were not "natural sounding" as the system merely strings together sounds for the ASCII characters interpreted. Nevertheless those words were still recognizable and became more so with continuing use.

(2) The system is also able to accept phonetically spelt words or phrases (with introduced inflections, pitch etc.) to enable "natural sounds to be output. This aspect has not yet been exploited.

(3) In addition to prompt field repeating, the voice response unit can be sent data from within a program or via any terminal on the system for the introduction of additional messages.

(4) In the configuration provided during the benchmark period, this aspect of the system was not considered to be of sufficient use to improve stereoplotter throughput. Further experience and exploitation of the voice response system may lead to more efficient usage, however the prompt field reporting is sufficient for all textual verification requirements identified to date.

d. Voice Recognition.

(1) A voice recognition system was provided which enabled programming of user specified feature selections and commands to be acceptable by voice input via a headset mounted microphone. A menu of voice utterances is designed, which can be networked to enforce a specific window (of following utterances) to be open for recognition following acceptance of any single utterance. During the benchmark period a number of different menus were designed, implemented and tested.

(2) The training for a menu of 50 utterances was shown to require 10 minutes of operator time. Immediately following the training session a test of recognition of those utterances gave a good result. The recognition factor dropped considerably during a days work to an estimated 50% hit rate. Although lack of experience with the unit can account for some failure, operator fatigue is the main factor in the drop in recognition rate. This could be relieved, to some extent, by re-training during a digitizing session (for single utterances or the complete menu).

(3) The basis of recognition is hardware dependant and advice received from the contractors is that the same operator will not be able to transfer his trained voice "template" from one unit to another if he were assigned to another stereoplotter.

(4) The traditional way of plotting topographic detail, has been one of continuous capture of one feature type at a time. Such a methodology enforces a logical data capture pattern that is least likely to omit features. In the digital environment this means that feature selections are not changed continually but rather that one selection is followed by a good deal of plotting (in most cases). The only area where many different features are likely is where cultural detail is involved.

(5) A situation in which the operator needed to be looking at his stereomodel at the same time as selecting a new command or feature type would be efficiently performed by voice input. Such a requirement may be one in which a number of different join commands are available for linear features. Since joins are requested continually during digitizing it would be prohibitive to require menu selection for such actions.

19. The result of the initial development benchmarking period was that model throughput times reduced considerably with each re-plot. The improving times were due to a number of factors, the most important of which were operator familiarity with the system and with the model area. Because of these factors, and the rate of development of the workstation, it was not possible to quantify the amount of saving attributable to individual introduced enhancements.

20. To achieve a true comparison of the developed stereoplotter workstation against throughput in the AUTOMAP 1 system a final series of tests was developed. The operator had, by that stage, replotted the same model four times and it is believed that his familiarity with the plotting requirement was not likely to change with further re-plots. To this end the same model was re-plotted on AUTOMAP 1 and on what was considered to be the most efficient AUTOMAP 2 configuration developed to date. Following the AUTOMAP 1 re-plot, that "unclean" data was edited both internally and externally to ensure that the data was to cartographic standards. (External editing involves joining to adjacent models.) From these figures, comparisons of plot time and total time required to produce clean model data were calculated.

21. The most efficient AUTOMAP 2 configuration developed to that stage included interactive graphics and image superimposition with command and feature selection via tbe menu system. The voice recognition/response system was not considered to be developed to an efficiency level that would result in increased model throughput. The range of commands available at that stage did not include those that were earlier described as best implemented by voice recognition techniques and the recognition rate (compared to menu selection) would have led to a slower selection procedure at that stage of development. A tracing stand button pad, although available for command assignment, was not used during the re-plot test, and all command and feature selection was via the menu system.

22. The plotting time experienced in each re-plot is tabulated in

Table 6. As can be seen, the same model was plotted on the AUTOMAP 2 system in 78% of the time required on the AUTOMAP 1 system, a saving of 3 hours at the stereoplotter.

	Automap 1 - Keyboard + Pencil Plot	Automap 2 - Interactive Graphics + Menu + Image Superimposition	Automap 2 Plot Time Improvement (% of Automap 1)
Cultural	2.67	1.90	29
Drainage	4.42	4.22	5
Relief	4.67	3.07	34
Vegetation	2.17	1.68	23
All	13.93	10.87	22

TABLE 6 - FINAL BENCHMARK RESULTS (HOURS)
(Excluding Data "Cleaning" Operations)

23. The model throughput time has improved considerably more than the 22% saved in plot time when other factors are taken into account. The AUTOMAP 1 stereoplotter station cannot be assigned a new model before a final hardcopy plot is produced to verify data capture. This entails an extra 1.33 hours (average). The data then produced must be edited to correct line joins to "visual" cartographic requirements and joins must be made to adjoining model data. In AUTOMAP 1, both are attempted at the stereoplotter with the aid of a pencil plot, but are digitally corrected/edited following amalgamation into a 1:25 000 data base page. Both actions have been simulated to provide detail of the extra time required to clean up that data. Internal joins were achieved by editing the model area on a 1:25 000 verification plot (the plot time itself was not included). External join time was estimated by counting the number of joins required, dividing by two (since on average only two sides will require joins), and multiplying by the average time for the join operation (calculated by a separate study over a large number of such operations and found to be 40 seconds per join). Times calculated were 1.5 hours and 0.83 hours respectively. This then represents 13% of the total model requirements. Actual production timings, reduced to percentages and tabulated in Table 4, suggest that 30% is a more realistic requirement but the former figure (13%) will be used as a minimum saving.

24. Table 7 tabulates the results of comparative calculations and shows that the AUTOMAP 2 system can produce clean model data in not more than 62% of the time required on the AUTOMAP 1 system.

	Automap 1	Automap 2	Comparative Throughput
Model Plotting Time	13.93	10.87	78%
Verification Hard Copy	1.33	No longer required.	
Internal Linework Edit	1.50	Performed within plot time.	
External Linework Joins	0.83	Performed within Plot time.	
Total Model Requirement	17.59	10.87	62%

TABLE 7 - BENCHMARK RESULTS EXTENDED TO "CLEAN" MODEL DATA (HOURS)

AMORTIZATION CALCULATIONS

25. Table 8 tabulates the monetary figures used in amortization calculations in conjunction with the results of this benchmark.

Benchmarked Configuration	Capital Cost	Operating Cost (1)	Comments
Wild B8 + Environment	N/A	$12 000	Fixed costs
Operator	N/A	$20 000	1 Man Year
Interactive Graphics	$28 000	$ 2 800	
Graphics Superimposition	$ 6 000	$ 400	

Note 1. $AUST per annum.

TABLE 8 - MONETARY FIGURES FOR AMORTIZATION CALCULATIONS

26. Where the total time taken to produce clean data is taken into account, the amortization period for the benchmark configuration is calculated as follows:

$$\text{Saving per year} = .38 \times (\$32\ 000) - \$3\ 200 = \$8\ 960$$

$$\text{Amortization Period} = \frac{\$34\ 000}{\$\ 8\ 960} = 3.8 \text{ years}$$

CONCLUSIONS

27. The benchmark test period has led to the development of a work-station configuration that will produce clean model data in a maximum of 62% of the time required for that same quality of data produced on the AUTOMAP 1 system; representing a 38% increase in productivity.

28. The enhancement costs of the benchmarked stereoplotter workstation configuration can be amortized in a maximum of 3.8 years.

STATUS OF AUTO CARTO AT THE DEFENSE MAPPING AGENCY

F. C. Green
Headquarters, Defense Mapping Agency
Washington, D.C. 20305

ABSTRACT

During the last year, DMA has focused on increased utilization of auto carto techniques and processes in production. Although significant auto carto capabilities have existed in the past, a combination of a balancing of production emphasis and software development is now providing increased automation in map and chart production.

Discussed are developments in exploiting digital product data for maps and charts and the family of products concept. Applications of scale conversion through digital processes using existing repromat source for smaller scale maps are reviewed. New applications of auto carto scheduled for integration into production in the near future are also discussed.

BACKGROUND

The use of automation in the cartographic processes of compilation, revision and color separation is not new to DMA. In the early and mid 1970's, several different systems were used in the production of maps, aeronautical and nautical charts. Key among these systems was the Semi Automated Cartographic System (SACARTS). It provided for digitizing compilation manuscripts, performing editing of the digitized data as well as other functions inherent in the development of symbolized map data, and producing color separation plates. Also of significance was the Linear Input System, or LIS, a manual digitizing system which provided DMA's primary digitizing capability for many years.

During the late 1970's and early 1980's, the DMA emphasis was on producing digital data products for weapon systems. This emphasis was, to a certain extent, at the expense of auto carto. However, during this period, several significant things happened. Computer technology, such as higher density storage media required to support auto carto in the DMA environment, became available. It also became apparent that the possibility existed for utilizing a significant portion of the vast amount of digital product data for charting purposes. Raster based systems became available for more efficient digitizing. Adding to this has been a shifting of production emphasis that has allowed increased focus on auto carto development and implementation. All of these actions have placed auto carto in a significant role in the current and future DMA production plans.

EXPLOITATION OF DIGITAL DATA

DMA has a large amount of digital data which has been produced to support weapon systems and simulators for flight training. This data takes many formats but two of the most common are Digital Terrain Elevation Data (DTED) and Digital Feature Analysis Data (DFAD). Both products are part of the Digital Landmass System (DLMS). DTED consists of matrices of terrain elevations. DFAD consists of data representing point, line and area features, including positioning and descriptive data.

The use of DTED in chart production has been an early application of auto carto to exploit digital data. Software has been developed to plot contours from the DTED data base at selected intervals and at different projections. This process has been integrated into the production of 1:200,000 scale aeronautical charts and is now being tested for other scales.

The indications from production of the first sheets are that a resource savings of about 10-15% is common in the use of DTED over previous manual compilation methods. As production procedures are refined, it is expected that the savings will increase.

The use of DFAD for charting purposes has also been achieved in prototype production. Aeronautical charts at the 1:200,000 scale depict the radar significance of planimetric features. Previously, compilation of these Radar Significant Analysis Codes (RSAC) has been a manual operation. Software has now been developed to use the DFAD for RSAC generation, and a prototype sheet has been evaluated favorably by the user community. Some problems exist, however. As an example, the DFAD does not contain all of the RSAC features which the map specifications require. Small bridges which are radar significant are an example. This situation is expected to cause a reconciliation of the chart specification and the DFAD specification.

The use of these digital terrain and feature data bases for maps and charts shows substantial promise. It forms the basis for a "family of products" concept, an approach which, when fully developed, would provide for multi-product data bases satisfying DMA digital and map/chart product requirements. Incumbent in this approach and in the initial production described above is increased compatibility of digital and map/chart products. Another demonstrated benefit has been the resource savings and compression of pipeline time. The problems of matching digital capabilities with traditional map/chart specifications as well as matching map/chart production schedules with digital production so as to exploit the digital data base for maps and charts still remain to be solved.

AUTOMATED TECHNIQUES USING CARTOGRAPHIC SOURCE

Another area in which DMA is now in prototype production is scale conversion for the production of smaller scale maps/charts from existing repromat and/or digital data. The first sheet to be produced in this manner will be at the 1:100,000 scale using 1:50,000 scale repromat as the source. Digitization of existing film negatives (with the exception of the contour plate) is being accomplished using the SCITEX Response 250, a color raster/graphic edit system with laser plotter. For the prototype, DTED was available (at the proper refinement level) and contours will be produced at the smaller scale from the DTED data base. For future sheets, if DTED is not available, the existing 1:50,000 contour plates would be digitized on the Automatic Graphic Digitizing System (AGDS), a raster scanning, vectorizing and editing system. After processing on the AGDS, digitized contour data would be converted to DTED, and then plotted at the smaller scale. The digitized planimetric/hydrographic data from the four 1:50,000 source sheets are panelled and merged on the SCITEX system and reduced to 1:100,000 scale. Selection of features to be portrayed at 1:100,000 scale, feature generalization, and symbolized feature displacement are performed interactively on the SCITEX.

Substantial work remains to be done, however, in that software to filter out features not required for the smaller scale, as well as feature generalization and displacement are to be developed. Once again, this is an application of auto carto where evaluation of existing map/chart specifications, based on manual compilation, might result in a specification revision for a closer match with auto carto capabilities.

The application of auto carto techniques also is being effective in the production of nautical charts. These charts typically are at 1:75,000 scale and depict hydrographic as well as topographic and planimetric features. Imagery used for the topography and shoreline is compiled on the AS11 analytical stereoplotter and the resultant manuscript is scanned on the SCITEX, producing a preliminary color separate. The Advanced Cartographic Data Digitizing Systems (ACDDS), a vector formatted digitization and compilation system, is used for compilation of the hydrographic data using the shoreline color separate discussed above for control. The SCITEX was also used for preparation of text and area feature guides. The CRT printhead plotter, a high speed vector formatted plotter capable of producing reproducible quality film positives, was used to produce the symbolized color separates of the hydrographic data and the SCITEX plotter to produce separates of the topography, text and open window negatives. Type placement was originally done on the SCITEX; however, experience has shown this to be a very labor intensive operation and software has been developed to do type placement on the ACDDS with output to the CRT printhead plotter.

Work remains to be done in this application of auto carto for nautical charts but it has already been incorporated into the production process with a savings of approximately 25% in resources and 33% in pipeline time. A data base (in a format to support nautical charts) has been developed and digital data from the compilation of each chart is entered into the data base. This data base promises even greater savings and faster response time for revisions of charts included in the data base.

AUTOMATION ENVIRONMENT

There are a number of related areas which have significance for auto carto development within DMA.

- DMA is well into an upgrade of its large scale scientific and technical computers. This capacity will provide support as necessary, for the additional automation requirements of auto carto.

- DMA is nearing finalization of a standard linear format for digital cartographic feature data which will be used by applicable auto carto systems and will facilitate inter system transfer of data.

- DMA is developing an interactive, networked data base system. This system, when in place, will facilitate the access and utilization of MC&G data for cartographic applications.

- A high speed (50 million bits/second) local area network is under development at each of the Production Centers. This capacity will facilitate the inter system transfer of digital data in auto carto production and will support the data base environment mentioned above.

Collectively these developments are putting into place a production environment that fosters the efficient generation of auto carto products, storage and retrieval of data elements and promotes interproduct commonality and compatibility.

DEVELOPMENT ACTIVITIES

DMA is laying the groundwork for a much broader auto carto program in future years through the development/acquisition of a number of systems and hardware items. These include:

- Computer Assisted Photo Interpretation System (CAPI), which will extract feature data to support digital product programs and auto carto.

- Cartographic Compilation Revision System (CCRS), which will provide the capability to use photographic, cartographic, or digital source for compilation, recompilation, and revision of map/chart features.

- o Advanced Edit System (AES), which will provide the capability to edit and plot symbolized digital data in the preparation of color separates.
- o Clustered Processing System (CPS), which is designed to mosaic, merge, and validate planimetric digital data to meet digital product requirements as well as support map/chart development.
- o Laser Platemaker, which will use digital data to produce a pressplate.
- o Compilation Revision Edit System (CRES), which will provide an interactive tagging/digitizing capability to edit and tag selective features for map/chart revision.

Integration of these systems into the production environment will occur principally in the 1983-1986 time frame. Implementation will be accompanied by development of a substantial amount of software to ensure interoperability of these and existing systems for auto carto.

SUMMARY

In the past, DMA has had the potential for development of a substantial auto carto program, but because of emphasis on production of digital products and other considerations, a sustained significant production capability has only recently been established. There is strong incentive to continue development of auto carto, not only because of the traditional efficiencies of auto carto, but because of the potential to exploit the large digital product data bases and to increase compatibility of digital and map/chart products. Finally, an auto carto technology base exists in DMA now with plans and a commitment for expansion and development for the future.

UNE BASE DE DONNEES RELIEF A L'IGN FRANCE

par D. PIQUET-PELLORCE et A. BERNARD

Institut Géographique National
2 Avenue Pasteur
94160 Saint-Mandé - FRANCE

RESUME

Dans le courant de l'année 1984, l'IGN aura terminé la numérisation du relief national à moyenne échelle. Ces données représenteront une quantité d'information de 1,5 milliard d'octets (environ 200 millions de points saisis en 3 000 unités).

Il était donc temps de mettre au point un système permettant de gérer ces données et constituant un outil pratique pour les divers utilisateurs : rectification d'images spatiales, confection d'orthophotographies, cartes de pente, d'ensoleillement, d'intervisibilité, vues perspectives, estompage, courbes de niveau généralisées.

Nous avons choisi une base de données "en ligne" comprenant les éléments saisis (courbes de niveau et points cotés), en coordonnées géographiques.

Par suite, nous avone défini un logiciel qui permettra :

a) la gestion de la base : sauvegarde des informations, obtention de renseignements qualitatif et quantitatif sur le contenu de la base, correction et mise à jour des donnáes ; notons que ce dernier point sera très important pendant les deux ans que durera la montée en charge de la basa, essentiellement à cause des raccords entre unités de saisie qu'il faudra assurer.

b) l'utilisation de la base : fourniture de dessins d'une zone quelconque, à l'échelle de base ou à une échelle dérivée, calcul de modèles numériques de terrain à pas et orientation quelconques ainsi que de produits dérivés, dans un délai rapide (au plus une demi-journée).

Pour l'architecture de la base, nous avons retenu le principe d'un espace continu sur toute la France, divisé en pavés géographiques de même taille (environ 1/4 km2), les éléments étant chaînés d'un pavé à l'autre. Cette structure de données améliore l'accès à l'information tant pour les mises à jour de la base que pour les calculs de modàlea numériques.

Ce logiciel sera opérationnel à la fin de l'année 1983, et dès le début de 1984, nous pourrons commencar à changer la base et mettre les pramièrea données à la disposition des utilisateurs.

HISTORIQUE DE LA NUMERISATION

Dans la courant de l'année 1984, s'achèvera la numérisation du relief national à moyenne échelle. Quel long percours ! qui se souvient à l'I.G.N. de la première table à numériser en 1970, de la première planche de courbes de niveau numérisée, des premières numérisations photogrammétriques sur ruban perforé, des premières batailles informatiques pour dominer ces fichiars si importents pour l'époque ? Il fallut plus de cinq ans d'études pour voir et aborder les applications, pour essayer de maîtriser le contrôle des données avant d'envisager une numérisation régulière du relief. En 1976 l'IGN se dotait de cinq tables de numérisation GRADICON avec sortie sur bande magnétique et commençait la saisie systématique à partir des planches de la carte de base au 1:25 000. Parallèlement il étudiait avec une entreprise française la fabrication d'un scanneur en remplaçant sur une table traçante l'ensemble de dessin par un système de lecture électronique à partir d'une barette de photodiodes. Après quelques difficultés, en 1978 il était confronté à cette matrice de 22 millions de bits chacun représentant un point de $(0,1\ mm)^2$ et devait faire des prouesses pour dominer les algorithmes permettant de retrouver les courbes et compatibles avec la puissance de l'IBM 370/135 qui servait alors d'ordinateur central pour tout l'institut. En 1980 on décidait d'abandonner la numérisation des planches de courbes pour les zones très accidentées (Alpes, Pyrénées, Massif Central) et de couvrir ces régions par numérisation photogrammétrique à partir de prises de vues à 1:60 000 . Enfin,en 1981,après avoir maîtrisé le passage du mode "matriciel" au mode "vecteur" sur ordinateur VAX 11/780 et processeur de tableau AP120B en moins d'une heure pour une matrice de 100 millions de points, l'IGN est passé au contrôle interactif graphique avec console de visualisation couleur.

Ainsi, il aura fallu environ 20 personnes durant 8 ans, d'incessantes améliorations, une recherche constante de la fiabilité pour acquérir environ 150 à 200 millions de points saisis représentant essentiellement des tronçons de courbes de niveau.

RECENSEMENT DES BESOINS ET DES APPLICATIONS

Plus d'un milliard d'octets pourquoi faire ? Bien sûr, l'un des premiers buts est d'obtenir des modèles numériques de terrain (M.N.T.). Dans ce domaine la demande ne cesse de croître tant à l'extérieur de l'Institut où les études et les applications sont de plus en plus nombreuses et surtout croissent en volume. Si les cartes de pentes ou d'ensoleillement restent encore des demandes occasionnelles et les cartes d'intervisibilité l'objet essentiellement d'études, la confection d'orthophotographies croit régulièrement et la rectification d'images spatiales doit à partir de 1985 tendre vers une production régulière d'une scène SPOT par jour. Les applications cartographiques se révèlent très exigeantes à partir de M.N.T. pour l'estompage, alors que la généralisation cartographique pour le 1:100 000 et les échelles plus petites semble devoir s'appuyer tantôt sur le lissage de courbes tantôt sur le filage de courbes dans un M.N.T..

De fait beaucoup d'applications n'avaient pas jusqu'alors été sérieusement développées par manque de perspectives à court terme pour avoir les données. Maintenant que la numérisation s'achève,les applications sont réellement en train d'exploser et le besoin se fait sentir de mettre en place les moyens pour répondre,d'une part aux nécessités d'études, d'autre part à des grandes productions assez régulières (rectification d'images spatiales, confection d'orthophotographies, utilisations cartographiques telles que courbes de niveau généralisées et estompage).

Des études de faisabilité,des expériences et des démonstrations il faut passer à la production. La première condition de ce passage était d'abord l'existence des données couvrant le territoire national, les suivantes sont la fiabilité et la disponibilité réelle de ces données. Pour cela il est grand temps de passer des quelques 3 000 fichiers non parfaitement homogènes et répartis sur près de 3 00 bandes magnétiques à un ensemble informatique cohérent permettant à tous les usagers d'accéder simplement et rapidement aux données et comportant les outils essentiels pour mettre en oeuvre les applications de grande exploitation.

Pouvoir traiter chaque jour l'accès à des zones de plus de

60 km sur 60 km avec une orientation quelconque et produire une matrice au pas de 50 m, soit environ 1 200 sur 1 200 points pour rectifier une image "SPOT", pouvoir accéder à une zone de 40 km sur 60 km et en zone montagneuse sortir un modèle numérique au pas de 25 mètres pour produire une planche d'estompage et une planche de courbes de niveau généralisée, pouvoir dans l'heure produire le M.N.T. au pas de 20 m ou 25 m pour une orthophotographie, voilà les vraies raisons pour créer une banque de données du relief.

ARCHITECTURE DE LA BASE

Nous avons tenté de définir une architecture qui optimise les accès aux informations et leur traitement, pour toutes les utilisations envisagées. Nos réflexions ont porté successivement sur le choix des éléments à conserver dans la base, le choix des coordonnées, l'organisation des données et enfin, le choix du support d'archivage.

1 - Choix des éléments constitutifs de la base :

Les deux possibilités que nous avons envisagées sont l'échantillonnage en Z, c'est-à-dire les courbes de niveau et l'échantillonnage en xy, soit les modèles numériques de terrain (M.N.T.). Le premier présente l'avantage de rendre de façon plus précise la forme du terrain (équidistance variable selon le relief avec présence éventuelle de courbes intercalaires), et d'être adapté à la cartographie en courbes de niveau. Le second est mieux adapté en particulier aux problèmes de localisation et aux traitements informatiques en général, ce qui se traduit par le fait que l'élaboration de la plupart des produits dérivés du relief numérique nécessite le calcul préalable d'un M.N.T..Il semblerait donc séduisant d'archiver un tel modèle couvrant la totalité du territoire, de manière à éviter des calculs nombreux. Néanmoins, il faut garder à l'esprit le fait que, quelle que soit la méthode employée à l'I.G.N.F, restitution numérique des photographies aériennes ou numérisation de la carte, les éléments saisis ont été les courbes de niveau et les points cotés : il est donc indispensable de conserver ces éléments initiaux, tout échantillonnage des données entraînant nécessairement une perte d'information, si minime soit-elle. De plus, les problèmes posés par l'archivage, la gestion et l'utilisation d'un modèle numérique ne sont pas négligeables :

- le calcul de M.N.T. dans une projection, à un pas ou selon une oriantation différente de ceux du modèle de base serait assez coûteux, tout en donnant un résultat moins précis qu'en le recalculant à partir des éléments initiaux, courbes de niveau et points cotés.
- Ce modèle, pour satisfaire l'ensemble des utilisateurs, devrait être à un pas assez fin (20 mètres), d'où un volume de données important, environ 2,5 gigaoctets, qui s'ajouteraient aux 1,5 gigaoctets nécessaires aux courbes de niveau.
- La gestion de la base, comprenant une donnée unique, le relief, sous deux formes différentes, se trouverait très nettement compliquée.

Il nous est finalement apparu que ces inconvénients l'emportaient sur l'avantage principal qui aurait été d'éviter le calcul d'un M.N.T. dans un grand nombre de cas.

2 - Choix des coordonnées :

Le choix des courbes de niveau et des points cotés comme seuls éléments constitutifs de la base étant ainsi arrêté , il nous a fallu opter pour un système de coordonnées. Le système Lambert en quatre zones, utilisé pas l'I.G.N.F. pour le cartographie aux moyennes échelles, et dans lequel évidemment nous avons saisi les données, semblait tout indiqué. Malheureusement, il existe des discontinuités au passage d'une zone à l'autre, si bien que nous avons préféré utiliser les coordonnées géographiques : nous obtenons ainsi une base en une carte au lieu d'une base en quatre cartes, ce qui nous aurait posé des problèmes informatiques superflus. Cette décision a été confortée par le fait que les transformations de coordonnées géographiques en projection sont très peu onéreuses.

3 - Organisation interne des données :

Un soin tout particulier a été apporté au choix de cette organisation qui conditionne la souplesse d'utilisation de la base; le but poursuivi ici est la recherche d'un accès optimal aux informations, tout en les maintenant dans un volume raisonnable. La difficulté réside dans le fait qu'il faut conserver une structure en vecteurs chaînés et points isolés (les courbes de niveau peuvent être assimilées à des lignes brisées), indispensable pour le dessin cartographique des courbes de niveau, ainsi que pour les mises à jour et les corrections, tout en favorisant l'accès localisé aux données, ce qui ferait pencherpour une structure de type matriciel. Cette contradiction a été résolue par le découpage du territoire en pavés géographiques de petite taille (0,8 X 0,4 cgr, soit environ 550 X 400 mètres), qui sont les éléments d'une grande matrice couvrant la France, les données demeurant sous une forme vectorielle (identifieurs de courbes, points de courbes, points cotés) à l'intérieur de chaque pavé ; bien entendu, ces pavés sont regroupés en fichiers (20 x 40 cgr, c'est à dire une feuille au 1:50 000 normalisé) mais ceux-ci ne sont que des moyens destinés à faciliter la gestion du disque par le système, et jouent un rôle très secondaire. L'existence de ces pavés assure l'accès localisé (un pavé compte en moyenne 80 points, ce nombre pouvant monter, dans certains cas exceptionnels, jusqu'à 500 points), mais la nécessité d'une véritable structure vectorielle exige la constitution d'un système de chaînage entre courbes de niveau d'un pavé à l'autre : à chaque élément totalement inclus dans un pavé, sont associés deux pointeurs vers les deux éléments le précédant et le suivant dans les pavés adjacents, l'un ou l'autre de ces pointeurs (ou les deux) pouvant être absent dans le cas où on a affaire à une extrémité de courbe. Nous pouvons d'une certaine manière considérer que nous avons affaire, à un niveau macroscopique, à un M.N.T. sous forme de grille régulière, chaque élément de la grille étant un pavé (unité d'accès à la base), lui même étant, au niveau microscopique, une carte du relief en format vecteur.

Nous pensons avoir ainsi atteint nos objectifs :

- optimisation de l'accès aux informations contenues dans une zone quelconque : il suffit d'aller chercher les pavés concernés, la plupart l'étant entièrement ; pour les autres, il suffit de séparer les points utiles des points extérieurs à la zone considérée, ce qui ne saurait être long lorsque l'on sait qu'il n'y a que 80 points par pavé.

- conservation de la possibilité d'accès aux courbes de niveau par le chaînage expliqué plus haut ; ceci est indispensable pour le dessin cartographique sur traceurs vectoriels, et pour le logiciel de corrections. En effet, dans ce cas, le système doit considérer comme un élément unique ce qui apparaît comme tel à l'opérateur sur une console graphique, ou sur un dessin de contrôle. Il est même possible de suivre une courbe d'un bout à l'autre du territoire, sans discontinuité aucune pour l'utilisateur. Nous pouvons considérer que ce système, au prix d'une perte de place minime (les points occupent 78 % du volume des informations, le pavage 16 % et le chaînage 6 %) fournit à l'utilisateur un moyen d'accès extrêmement rapide aux données, dès lors qu'elles sont limitées dans une zone géographique déterminée et ce, malgré leur morcellement en une mosaïque de plus de 2 500 000 pavés. De plus, la continuité absolue des courbes de niveau est assurée sur tout le territoire au seul prix d'une légère dégradation des performances uniquement fonction de la taille de la zone interrogée.

4 - Choix du support :

On peut songer d'abord à un archivage sur bandes magnétiques, a priori plus économique. Néanmoins, celui-ci présente dans notre cas un certain nombre d'inconvénients. D'une part, il est prévu, nous le verrons plus loin, d'assurer un grand nombre de corrections pendant deux ans, d'autre part, il faudrait répartir les fichiers sur différentes bandes. Etant données la grande taille des zones demandées par la rectification d'images spatiales (au moins 60 x 60 Km) et la nature des zones demandées par les corrections à cheval sur plusieurs fichiers (situés sur les bandes magnétiques différentes), les temps de réponse seraient très pénalisants. Un archivage sur support à accès direct s'avère nécessaire. Il est raisonnable de l'envisager, dans l'état technologique actuel, sur disques magnétiques. (quantité d'information prévue : 1,5 gigaoctets).

LOGICIEL

Nous distinguerons ici les programmes de gestion de la base et les programmes d'utilisation qui fournissent soit un produit "brut" (M.N.T., fichiers de C.N.N.) soit un produit plus élaboré (carte de pente, d'intervisibilité, dessin des courbes généralisées ou non).

1 - Gestion de la base :

Les fonctions assurées sont l'insertion de nouvelles données, l'extraction, le remplacement de données anciennes, la gestion de l'état de la base, les corrections et mises à jour ainsi que les transformations de coordonnées. Les trois premiers points n'appellent pas de remarques particulières si ce n'est que ces opérations sont faites par nombre entier de pavés.

- Gestion de l'état de la base

Sont enregistrés, pour chaque fichier (feuille au 1:50 000 normalisée) la méthode de numérisation, les dates de saisie des informations et des corrections éventuelles, des indications sur la qualité des données, que ce soit en vue de l'édition cartographique des C.N.N. ou de l'utilisation à partir d'un M.N.T., ainsi que des indications quantitatives (altitudes extrêmes, nombre de points, équidistance ...) Sont prévues, des sorties graphiques et des états imprimés de ces informations.

- Correction et mise à jour

C'est la partie la plus volumineuse de ce logiciel de gestion ; il est à noter qu'elle sera utilisée intensivement durant la montée en charge de la base : il est en effet prévu d'y affecter une dizaine de personnes pendant deux ans. Les corrections concernent les erreurs dues à la numérisation, mais surtout les raccords entre les unités de saisie. A l'issue de ce laps de temps, nous posséderons une base suffisamment fiable dont le rythme de mise à jour sera beaucoup plus lent, suivant la réfection de la carte de base. L'opérateur ayant indiqué la zone sur laquelle il désire travailler, le

système extrait les pavés concernés et constitue un fichier de travail qui sera modifié au cours des séances de correction ; une fois celles-ci terminées, il replace dans la base les pavés utilisés. Le programme de correction utilise une table à numériser, une console graphique couleur à balayage télévision, et une console alphanumérique, le tout connecté sur mini-ordinateur.
Le pavage défini plus haut est doublement bénéfique :
a) l'affichage sur la console graphique se fait par pavés entiers.
b) la désignation d'une courbe par pointé sur la table à numériser est extrêmement rapide (1 à 5 dixièmes de seconde), car la recherche ne se fait que dans un seul pavé. De plus, la souplesse d'agencement des pavés permet de se définir un fichier de travail centré sur une zone intéressante (par exemple une ligne de raccords).

- transformations de coordonnées

Celles-ci étant archivées en coordonnées géographiques, et la plupart des utilisateurs travaillant dans d'autres systèmes de coordonnées, les transformations correspondantes seront utilisées très fréquemment ; il est donc indispensable qu'elles soient rapides. C'est le cas grâce au pavage du territoire. En effet la taille des pavés est telle qu'à l'intérieur d'un groupe de plusieurs pavés, toutes ces transformations sont linéarisables, sans aucune perte de précision. Il s'ensuit que le coût informatique de ces transformations est très faibles (environ 15 000 points transformés par seconde).

2 - Utilisation de la base :

Une fois mis en place les outils permettant de gérer la base, il s'agit de fabriquer ceux grâce auxquels les divers utilisateurs pourront utiliser la base de manière sure et efficace ; ils pourront obtenir des produits brute, qui constituent le service minimum de la base, ou des produits finis, cartographiques ou numériques. Nous allons passer en revue tous les produits qui seront disponibles dès le démarrage de la base, étant entendu que la liste n'en est pas limitative, des développements étant toujours en cours à l'I.G.N.F. Précisons que pour tous les produits, le système de coordonnées est paramétrable, et que la zone de travail est définie par un polygone quelconque. Pour les sorties graphiques, nous pouvons actuellement utiliser,au choix, traceurs rapides (encres, pointes billes, feutres), imprimantes électrostatiques ou laser caméra. Les courbes de niveau seront,au gré de l'utilisateur, chaînées ou non, une équidistance différente de celle de la base pourra être spécifiée, un algorithme filtrant pourra éliminer un certain nombre de points jugés superflus.

Les M.N.T. seront le produit le plus demandé; il est donc particulièrement important de les fournir à un coût raisonnable avec la meilleure précision compatible avec les données. Les méthodes de calcul à l'I.G.N.F. comme dans beaucoup d'autres organismes, ont fait l'objet de nombreuses études et développements depuis maintenant plus de dix ans. La méthode que nous exposons ici est employée à l'I.G.N.F. depuis dix-huit mois en production.

On commence par déterminer les intersections des courbes de niveau avec les droites du maillage x = xi et y = yj. Sur chaque droite x = xi, on obtient une suite d'intersections que l'on classe par ordonnées croissantes, et à partir desquelles on interpole les altitudes aux noeuds du maillage. L'interpolation est linéaire ou cubique selon que les altitudes entourant le noeud sont distinctes ou non. De manière symétrique, on obtient les altitude aux noeuds à partir des intersections avec les droites y = yj. On adopte alors comme altitude du noeud l'altitude correspondant à la direction selon laquelle la pente est la plus forte,si elle existe, la moyenne des deux altitudes sinon. On introduit alors les points cotés en affecfectant leur altitude au noeud le plus proche. On obtient à ce moment un modèle approché. Le modèle définitif est calculé par la méthode de la grille élastique de M. de Masson d'Autume programmée sur processeur de tableau AP 120 B. Les valeurs Zij des altitudes aux noeuds sont les solutions au sens des moindres carrés du système linéaire suivant :

$$Z_{i-1,j} + 2\,Z_{i,j} + Z_{i+1,j} = 0$$
$$Z_{i,j-1} + 2Z_{i,j} + Z_{i,j+1} = 0$$
$Z_{i,j} = z_{i,j}$ avec un poids $P_{i,j}$

Les $z_{i,j}$ étant les altitudes approchées.

Le choix des $P_{i,j}$ est délicat ; celui qui a été fait s'est révélé satisfaisant: dans le cas où $z_{i,j}$ est l'altitude d'un point coté, $P_{i,j}$ à une valeur constante élevée, sinon, d étant la plus courte distance d'un noeud à une courbe de niveau suivant une des deux directions du réseau, $P_{i,j}$ est une fonction décroissante de d. Catte méthode, éprouvée en production, fournit des M.N.T. de qualité satisfaisante, avec lesquels on peut envisager des applications très exigeantes telles l'estompage et la généralisation. Les temps de calcul, avec les matériels dont nous disposons actuellement, VAX 11/780 de DEC et AP 120 B de FPS, sont de 10 à 15 minutes pour une grille de 300 sur 300 au pas de 50 mètres. Le pas étant fixé, ce temps est proportionnel au nombre de noeuds, par contre, si le nombre de noeuds est fixé, c'est une fonction décroissante du pas : en effat, plus le pas est faible, plus les $P_{i,j}$ sont faibles, donc plus la convergence est longue à atteindre.

Les produits dérivés de M.N.T. les plus courants sont les vues perspectives, les hypsomètries, les estompages et les cartes de pente, ces dernières n'étant d'ailleurs que des estompages particuliers avec un soleil au zénith.

Seuls parmi ceux-ci, les estompages risquent d'être produits en grand nombre pour les besoins internes à l'Institut. Un cas un peu particulier est celui de la généralisation des courbes de niveau. Deux méthodes totalement différentes, qui donnent des résultats très comparables, ont été développées. La première concerne le lissage des courbes de niveau, la seconde est une interpolation de ces mêmes courbes à travers un M.N.T. Les essais que nous avons effectués montrent que la première méthode s'avère meilleure dans les terrains très peu accidentés et où les courbes de niveau sont assez indépendantes les unes des autres et, par ce fait, peuvent être lissées indépendammant, alors que la seconde est nettement préférable dans les terrains accidentés, là où les courbes s'emboîtent et où le traitement doit être global pour respecter ces emboîtements.

Les résultats obtenus soutiennent, du point de vue de la qualité, la comparaison avec la généralisation traditionnelle ; de plus, les délais et les coûts sont de beaucoup inférieurs (une vingtaine d'heures d'opérateur contre 150 heures de dessinateur pour une feuille à 1:100 000), si bien que la production des planches d'altimétrie à 1:100 000 a commencé dès cette année par méthode automatique.

CONCLUSION

Dès la fin de 1983, avant même que la numérisation du relief français soit totalement terminée, nous disposerons des outils pour charger la base.

Dès le début de 1984, les corrections pourront commencer et les utilisateurs pourront faire calculer les premiers modèles numériques de terrain.

L'ensemble du logiciel que nous avons décrit sera intégré l'année prochaine et continuera à être développé. Nous avons la certitude que cette base va rendre opérationnelles les productions automatisées à partir du relief numérique déjà programmées : généralisation, estompage, orthophotographies, rectification d'images spatiales (niveau 3 de SPOT) et faciliter les études et développements de nouveaux produits dont on perçoit nettement le besoin grandissant. Elle est aussi le premier élément d'une base de données topographiques dans la constitution de laquelle l'I.G.N.F. s'engage résolument.

Bibliographie :

(1) A. Bernard - Digitization of relief data and exploitation of digital terrain model at I.G.N.F. actes d'Auto-carto V 1982

(2) G. de Masson d'Autume - Construction du Modèle numérique d'une surface par approximations successives. Application aux modèles numériques de terrain. Bulletin n° 71/72 de la société Française de Photogrammétrie et de Télédétection PARIS 1978.

FABRICATION DE BASES DE DONNÉES À RÉFÉRENCE SPATIALE A PARTIR DE DONNÉES NUMÉRIQUES DE LA CARTE DE BASE DU QUÉBEC À L'ÉCHELLE 1:1 000

Gervais Pellerin, a.g., M.Sc.
Société de Cartographie du Québec
1650, rue Louis-Jetté
Québec, QC Canada
G1S 2W3

ABSTRACT

Since September 1981, the Quebec Cartography Company was producing 1:1 000 photogrammetric digital maps for various ministries and companies using an interactive graphics design system. The Quebec Province digital mapping standards are used to build position and representation data files.

In the paper, we include a description of the data base schema defined specially for base maps and also a discussion of the task that «Tags» the non-graphic data to graphic elements.

The paper will finally set a procedure in which a representation data file is first translated into a cartographic data base system and, at the end, into different special geographic base information systems.

INTRODUCTION

La Société de Cartographie du Québec oeuvre dans le domaine de la cartographie assistée par ordinateur depuis septembre 1981 et exécute, avec l'aide d'un système graphique interactif de la compagnie INTERGRAPH, différents types de travaux pour le compte d'organismes publics ou parapublics et de compagnies privées.

Pour la fabrication de cartes numériques à grande échelle on applique les normes du service de la Cartographie du ministère de l'Energie et des Ressources du Québec pour la constitution des fichiers de captage et d'édition (échelle graphique 1:1 000).

Récemment, suite aux demandes accrues pour la création de base de données à référence spatiale où l'on associe les données littérales (numéro, nom de groupe, etc.) à leur représentation graphique (droite, courbe, symbole), une procédure a été développée pour créer un nouveau produit mieux adapté aux usages multiples que l'on appelle base de données cartographiques, qui contient l'information graphique et littérale de la carte numérisée.

FICHIER DE POSITION ET DE REPRÉSENTATION

Les données graphiques des cartes numériques à grande échelle (1:1 000) fabriquées suivant les normes du service de la Cartographie du ministère de l'Energie et des Ressources du Québec proviennent de:

- traduction (adaptation) de données existantes
- stéréorestituteur (digimètre 3 axes)
- table numérisante (digimètre 2 axes)
- entrée au clavier

Les principales caractéristiques de ces fichiers topographiques sont les suivantes:

- 63	niveaux	(overlay)
- 32	épaisseurs de trait	(thickness)
- 5	symbolisations	(line style)
- 256	couleurs	(color)
- 19	types d'éléments	(element type)

On retrouve sur les fichiers de position (captage) et sur les fichiers de représentation (édition) les types d'éléments graphiques suivants:

- éléments géométriques: (cercle, carré, rectangle, droite etc.)
- segments de droite, point par point
- courbes, point par point
- courbes avec points à intervalle régulier
- cellules et cellules répétitives
- textes

Les fichiers de position (captage) contiennent la position géographique des éléments cartographiques; tandis que les fichiers de représentation (édition) contiennent la symbolisation cartographique effectuée en vue d'un tracé sur un support quelconque, à une certaine échelle. Si la précision géométrique est essentielle pour l'usager, on doit alors utiliser les fichiers de position au lieu des fichiers de représentation.

L'analyse de la structure des fichiers ayant ces caractéristiques nous amène à la définition d'une cassification générale qui facilite l'association de données littérales à chacun des éléments graphiques.

CONCEPT GÉNÉRAL RETENU POUR L'ASSOCIATION

Les données numériques des fichiers de représentation, comme nous venons de le voir, sont déjà codifiées pour faciliter le traitement graphique. Les quatre paramètres suivants: niveaux d'information, largeur du trait, symbolisation de la ligne et type d'éléments vont nous permettre d'identifier chacun des éléments graphiques du fichier et de leur associer des données littérales.

L'optimisation de cette opération nous a amené à choisir le traitement par lot (batch processing) pour transformer le fichier original de représentation en un nouveau fichier contenant à la fois les données graphiques et littérales, que l'on appelle «base de données cartographiques». Les opérations sont effectuées en utilisant toutes les fonctions de base fournies par le système graphique; le mode interactif étant réservé pour la vérification et les modifications ultérieures.

Nous constatons à l'heure actuelle que l'approche la plus valable est la création d'une base de données cartographiques minimale dont la structure permettra d'accéder facilement aux différentes bases de données générales ou spécialisées. La classification des éléments suivant leur nature correspond en fait aux normes utilisées pour la représentation graphique (ex. bâtiments administratifs, bâtiments résidentiels, etc.). Comme il est important d'assurer un échange d'informations entre les différents usagers de données cartographiques on ajoute, si possible, les équivalences avec les normes canadiennes de stockage et d'échange d'informations topographiques numériques.

Comme on peut le voir à la figure 1, nous pouvons finalement modifier et adapter la base de données cartographiques aux différentes bases de données spécialisées qui sont propres à chaque usager (ex. municipalités, compagnies de services publics, etc.).

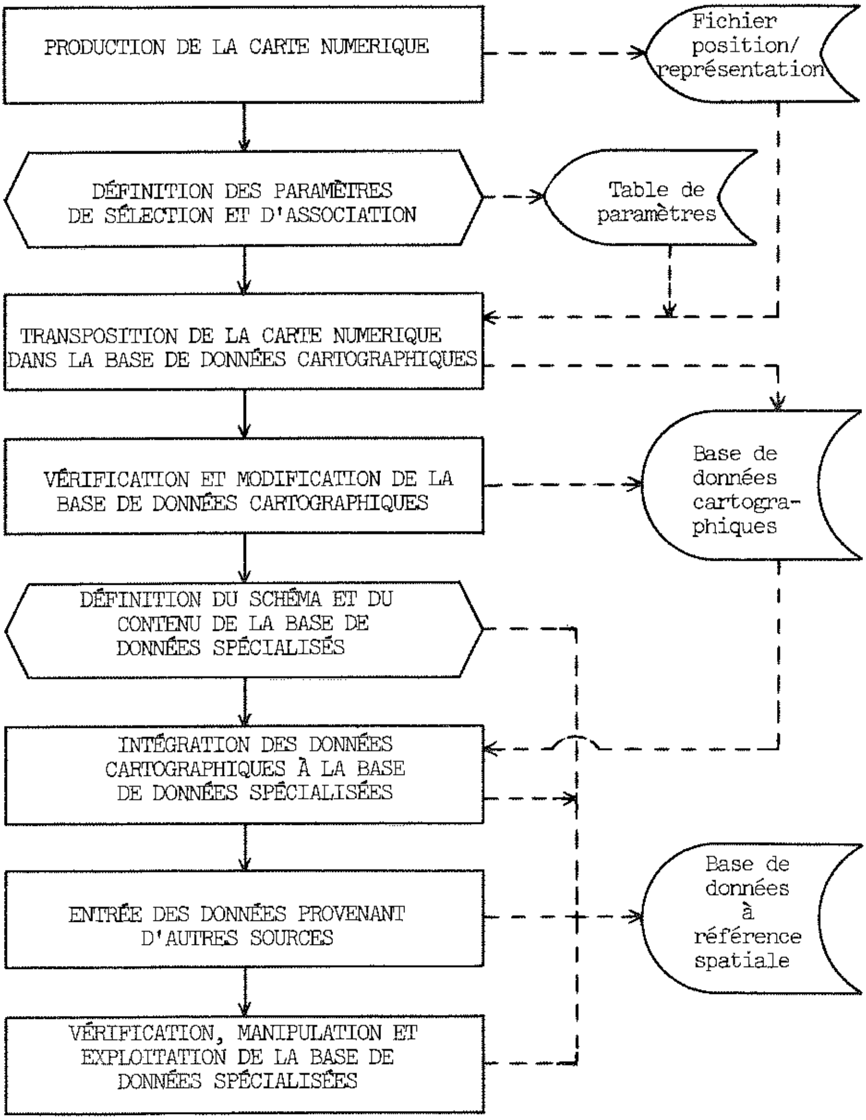


Figure 1. Composantes opérationnelles du processus

FABRICATION DE LA BASE DE DONNÉES CARTOGRAPHIQUES

Voici une description sommaire des étapes suivies pour réaliser la base de données cartographiques.

La premières étape consiste à élaborer le schéma de la base de données cartographiques. L'utilisation du système graphique interactif nous a permis d'analyser diverses solutions pour finalement en arriver à une structure où les éléments graphiques sont regroupés sous 27 classes ou groupes, selon leur nature. Une classe supplémentaire assure la codification des éléments non pertinents ou en erreur. On retrouve à la figure 2 la liste des classes (entités) de cette base de données cartographiques.

```
PR = 'QSO: (110,00)Q1000DB.LST'
DB = 'QSO: (110,11)Q1000DB.DBS'
NAME = 'Q1000'
PS = 'DB'
                              INDEX DES FICHIERS D'ENTITÉS

     1. AME-REL            : aménagement du relief
     2. AME-ROU-PRI        : aménagement routier privé
     3. AME-TOU            : aménagements touristiques
     4. BAT-ADM-COM        : bâtiments administratifs ou commerciaux
     5. BAT-ANN            : bâtiments annexés
     6. BAT-RES            : bâtiments résidentiels
     7. CON-REL-HYD        : constr. relatives à l'hydrographie
     8. CON-REL-VOI-COM    : constr. relatives aux voies de communication
     9. CON-URB-IND-AGR    : constr. urbaines, industrielles ou agricoles
    10. COU-NIV            : courbes de niveau
    11. DEM                : démarcation
    12. GEO-HYD            : géologie-hydrographie
    13. HYD                : hydrographie
    14. LIG-ELE            : lignes électriques
    15. LIM-ADM            : limites administratives
    16. LIM-TER            : limites terrestres
    17. PHO                : photogrammétrie
    18. POI-COT            : points cotés
    19. SER                : services
    20. SER-MUN            : services municipaux
    21. TER-DIV            : terrains divers
    22. TER-EXP            : terrains d'exploitation
    23. TER-REC            : terrains récréatifs
    24. TOU                : tours
    25. VEG                : végétation
    26. VOI-FER            : voies ferrées
    27. VOI-PUB            : voies publiques
    28. ZZZ                : autres, indéterminés

    SCHEMA DE LA BASE DE DONNÉES «Q1000DB.DBS»  VERSION 1.0  830707
         DÉVELOPPÉ PAR LA SOCIÉTÉ DE CARTOGRAPHIE DU QUÉBEC
```

Figure 2. Liste des entités de la base de données cartographiques

A l'étape suivante, on identifie les paramètres assignés à chacun des éléments graphiques contenus dans la carte numérique et on fabrique un fichier «table de paramètres». La codification jumelée à cette table repose sur les normes canadiennes en matière d'informations topographiques et sur une identification simplifiée des éléments et sous-éléments graphiques. La figure 3 nous montre les attributs associés à l'entité «bâtiments résidentiels».

ENTITÉ BÂTIMENTS RÉSIDENTIELS ET SES ATTRIBUTS

```
6.  BAT-RES       P=0             DF='QSO:(110,11)Q1000E06.ENT'

 .1  XREF         F=X;            ATTRIBUT DE RÉFÉRENCE
 .2  INDICATEUR   F=AN(1);        ÉTAT TEMPORAIRE (FLAG)
 .3  NUMERO-SCQ   F=I(99999);     NUMERO UTILITAIRE
 .4  NOM-QUE      F=C(6);         NOM (QUÉBEC)
 .5  NUM-QUE      F=C̄(5̄6);        NUMÉRO (QUÉBEC)
 .6  NOM-CAN      F=C̄(1̄06̄);       NOM (STANDARD CANADIEN)
 .7  NUM-CAN      F=C̄(1̄56̄);       NUMERO (STANDARD CANADIEN)
```

LISTE DES CODES POUR CETTE ENTITÉ

```
C6      MAX=11;              BAT-RES  .NOM-QUE

  1 = 'BALCON'
  2 = 'COUPE-FEU'
  3 = 'MAISON'
  4 = 'MAICON'
  5 = 'MAIRUI'
  6 = 'PERRON'

C56     MAX=11;              BAT-RES  .NUM-QUE

  1 = '23010'
  2 = '21010'
  3 = '21040'
  4 = '21020'
  5 = '21030'
  6 = '23020'

C106    MAX=11;              BAT-RES  .NOM-CAN

  1 = 'UNKNOWN'
  2 = 'UNKNOWN'
  3 = 'HOUSE'
  4 = 'BUILDING  (UNSPECIFIED)'
  5 = 'RUIN'
  6 = 'UNKNOWN'

C156    MAX=11;              BAT-RES  .NUM-CAN

  1 = '              '
  2 = '              '
  3 = 'BN 13700 000'
  4 = 'BR 03200 000'
  5 = 'BL 25200 000'
  6 = '              '
```

Figure 3. Partie de la base de données cartographiques

Une fois ces deux étapes réalisées, un logiciel effectue rapidement la correspondance entre les éléments graphiques et la table de paramètres puis greffe les attributs de la base de données cartographiques. On complète le traitement par une vérification en mode interactif.

MODIFICATION DE L'INFORMATION CARTOGRAPHIQUE ET TRANSFERT DANS LES BASES DE DONNÉES SPÉCIALISÉES

Une fois réalisée la constitution de la base de données cartographiques, nous mettons en marche le processus de fabrication des bases de données spécialisées avec l'arrivée de chaque projet particulier. A ce stade-ci, si la base n'existe pas, nous élaborons avec la collaboration des usagers, une structure qui répond parfaitement à leurs besoins. A la figure 4 nous trouvons la liste des entités d'un système d'information à référence spatiale.

```
PR = 'QSO:(110,11)SIRS.LST'
DB = 'QSO:(110,11)SIRS.DBS'
NAME = 'SIRS'
PW = 'DB'
```

INDEX DES FICHIERS D'ENTITÉS

1.	BORNE	: BORNE D'ARPENTAGE OU CADASTRALE
2.	LIGNE	: LIGNE D'ARPENTAGE OU CADASTRALE
3.	LOT	: ENTITÉ CADASTRALE
4.	BATI-PRINC	: BÂTIMENT PRINCIPAL
5.	MÉNAGE	: ENTITÉ FAMILLIALE
6.	BATI-SECON	: BÂTIMENT SECONDAIRE
7.	VOIRIE	: ROUTE, CHEMIN, ETC.
8.	SERVICE-PUBLIC	: EAU, ELECTRICITÉ, TÉLÉPHONE, ETC.
9.	ZONE	: ZONAGE URBAIN OU AGRICOLE

SCHÉMA DE LA BASE DE DONNÉES 'SIRS.DBS' VERSION 1.0 830707
DÉVELOPPÉ PAR LA SOCIÉTÉ DE CARTOGRAPHIE DU QUÉBEC

Figure 4. Liste des entités d'une base de données spécialisées: «SIRS»

Si par contre, la base spécialisée est déjà structurée, nous passons directement au transfert de l'information cartographique pertinente dans cette base de données spécialisées. Cette intégration est réalisée avec l'aide des logiciels existants du système graphique interactif. La figure 5 nous fait voir une partie des attributs d'une base de données spécialisées «SIRS» et un exemple de la codification employée.

Chacune des bases de données spécialisées nécessitent des opérations particulières et une bonne partie du travail d'intégration peut être exécutée en dehors du mode graphique interactif. Quelques fois, il est nécessaire d'utiliser une sortie graphique sur laquelle apparaît l'information littérale de la base de données cartographiques pour faciliter la localisation de l'élément graphique (référence spatiale) où se greffe cette information.

ENTITÉ BÂTIMENT PRINCIPAL ET SES ATTRIBUTS

4.	BATI-PRINC	P = 0	DF=QSO:(110,11)SIRSE04.ENT
.1	XREF	F = X	ATTRIBUT DE RÉFÉRENCE
.2	INDICATEUR	F = AN(1);	ETAT TEMPORAIRE (FLAG)
.3	NUMÉRO-SCQ	F = I(99999);	NUMÉRO UTILITAIRE
.4	REG-ADM	F = C(7);	RÉGION ADMINISTRATIVE
.5	NUMÉRO	F = I(999);	NUMÉRO DE DOSSIER
.6	LOT	F = AN(10);	LOT CADASTRAL
.7	CATÉGORIE	F = C(8);	CATÉGORIE DE BÂTIMENT
.8	CARACTÉRISTIQUE	F = C(9);	CARACTÉRISTIQUE DU BÂTIMENT
.9	IDENT	F = AN(20);	IDENTIFICATION
.10	DATE-CONSTR	F = I(999999);	DATE DE CONSTRUCTION
.11	QUALITÉ	F = C(10);	CODE DE QUALITÉ
.12	SURFACE	F = F;	SUPERFICIE DU BÂTIMENT (m carré)
.13	FONDATION	F = C(11);	TYPE DE FONDATION
.14	SER-EAU	F = C(12);	TYPE DE SERVICE D'EAU POTABLE
.15	SER-EAU-USÉE	F = C(13);	TYPE DE SERVICE D'EAU USÉE
.16	SER-ÉLEC	F = C(14);	TYPE DE SERVICE D'ÉLECTRICITÉ
.17	SER-TÉLÉ	F = C(14);	TYPE DE SERVICE DE TÉLÉPHONE
.18	CHAUFFAGE	F = C(16);	TYPE DE CHAUFFAGE
.19	ÉTAGE	F = C(15);	NOMBRE D'ÉTAGES
.20	CHAMBRE	F = I(99);	NOMBRE DE CHAMBRES
.21	BÂT-SECON	F = I(99);	NOMBRE DE BÂTIMENTS SECONDAIRES
.22	PROPRIÉTAIRE	F = AN(30);	NOM DU PROPRIÉTAIRE
.23	NUM-PROP	F = I(9999);	NUMÉRO PERSONNEL (PROPRIÉTAIRE)
.24	CAT-PROP	F = C(33);	CATÉGORIE DE PROPRIÉTÉ
.25	OCCUPANT	F = AN(30);	NOM DE L'OCCUPANT
.26	NUM-OCCU	F = I(9999);	NUMÉRO PERSONNEL (OCCUPANT)
.27	CAT-OCCU	F = C(17);	CATÉGORIE D'OCCUPATION
.28	INDIVIDU	F = I(999);	NOMBRE D'INDIVIDUS
.29	MÉNAGE	F = I(99);	NOMBRE DE MÉNAGES

LISTE D'UNE PARTIE DES CODES POUR L'UNITÉ BÂTIMENT PRINCIPAL

```
C15               MAX = 16           ;(BATI-PRINC) NOMBRE D'ÉTAGES
   1 = '1'
   2 = '1 1/2'
   3 = '2'
   4 = '2 1/2'
   5 = '3'
   6 = '3 1/2'
   7 = '4'
   8 = 'AUTRE'

C16               MAX = 20           ;(BATI-PRINC) TYPE DE CHAUFFAGE
   1 = 'BOIS'
   2 = 'MAZOUT (HUILE)'
   3 = 'ÉLECTRICITÉ'
   4 = 'AUCUN'
   5 = 'MAZ. + ELEC.'
   6 = 'MAZ. + BOIS'
   7 = 'ELEC. + BOIS'
   8 = 'AUTRE'
   9 = 'INCONNU'
```

Figure 5. Partie de la base de données spécialisées «SIRS»

LIMITE DE LA MÉTHODE EXPOSÉE

Un certain nombre de considérations techniques se dégagent de la présentation sommaire de la méthode. La constitution des fichiers de position et de représentation répond parfaitement aux normes de présentation graphique monochrome (cartographie numérique), mais lorsque l'on ajoute des données littérales à chacun des éléments graphiques, on fait face à un nombre plus ou moins grand de difficultés.

En premier lieu, l'automatisation du procédé se limite aux contraintes de liaison des éléments entre eux (chaînage) et à la multiplication des éléments graphiques pour couvrir une surface donnée. Dans ce dernier cas, il faut éliminer manuellement toutes les données non pertinentes et regrouper les éléments de la surface sous une seule occurence. La problème sera résolu un jour avec la carte couleur où la surface sera délimitée par un contour auquel l'on greffera l'information littérale.

Il est important de constater que les éléments graphiques sont codifiés suivant les normes canadiennes ce qui constitue un pas vers l'échange d'informations topographiques. L'organisation des données pour les bases appliquées est une étape longue et laborieuse, la méthode exposée ici facilite grandement l'étape de codification des données cartographiques utiles à chacune des bases de données appliquées.

Finalement, il nous faut analyser chaque application de façon bien particulière, car les éléments importants dans un cas ne le sont plus pour un autre.

CONCLUSION

Le développement d'une méthode permettant la fabrication de banques de données appliquées à partir de données numériques de la carte de base du Québec n'est rendu qu'au stade primaire. La présentation vise à sensibiliser les fabricants et usagers de produits cartographiques aux besoins grandissants de banques données à référence spatiale.

Si l'on veut s'assurer d'un avenir intéressant dans l'exploitation des fichiers de la carte de base du Québec, nous devons établir dès maintenant une procédure simple pour ajouter la constituante très importante: l'information littérale ou données non graphiques.

De toute évidence, nous aurons un outil essentiel au développement des systèmes d'information spécialisés à référence spatiale.

BIBLIOGRAPHIE

Bureau de la Cartographie et des Arpentages 1982, Vers un Système d'Information spatiale, Notes sur l'objectif et la stratégie, ministère de l'Energie et des Ressources du Québec.

Conseil canadien des Sciences géodésiques 1982, Standards for the Classification of Topographic Features, ministère de l'Energie, des Mines et des Ressources du Canada, section des Levés topographiques.

Gamache, A. 1980 Logiciel SGBD et Base de Données, les Presses de l'Université Laval, Québec.

Service de la Cartographie 1982, Normes pour la Constitution des Fichiers de Captage et d'Édition, ministère de l'Energie et des Ressources du Québec.

Société de Cartographie du Québec 1983, Développement d'un Système de Gestion à Référence spatiale, (Rapport pour E.M.R. Canada, section Cartographie foncière, Bureau régional du Québec).

THE FUTURE IN TERRAIN ELEVATION DATA PROCESSING AT THE DEFENSE MAPPING AGENCY

Mr. Robert B. Edelen
Defense Mapping Agency
Washington, D. C. 20315

ABSTRACT

Since 1968 the Defense Mapping Agency (DMA) has been actively engaged in the production of Digital Terrain Elevation Data; initially to support plastic relief mapping, then to support "lines of site" studies, and more recently to support advanced weapons systems, simulators and navigational devices. With the increasing variety of users for Digital Terrain Elevation Data (DTED), the requirement for greater quantity, accuracy and diversity of DTED has likewise increased. To help meet these challenging production requirements of the future, DMA has awarded a major contract for development of the Terrain Edit System/Elevation Matrix Processing System (TES/EMPS). The TES/EMPS will provide state-of-the-art mini computers, array processors and specially developed software to perform complex data transformation, interpolation, merging and discrepancy detection functions. The TES/EMPS work stations, combined with mini computers and graphics processors, will provide advanced graphic displays and editing cababilities to perform interactive error correction, panelling and validation functions. All work flow through the system will be managed by the sophisticated job and file management software. The TES/EMPS will be installed at DMA in December 1983 and will perform all processing of DTED collected from photographic and cartographic sources.

INTRODUCTION

A major product of DMA is the Digital Elevation Matrix (DEM). A DEM refers to any set of elevation data recorded on a uniform grid; and consists of data defining the corners of a rectangular array, the grid spacing, and an array of elevation values. AT DMA, DEMs are produced in a variety of formats, coordinate systems and grid spacings to satisfy an increasing number of digital data users.

For those users who require digital data in a geographic coordinate system, the DMA-Standard format is used. DMA-Standard DEMs are produced in "cells" covering 1° X 1°, 15' X 15' or 7.5' X 7.5' with data spaced at 3" X 3", 1" X 1" or 0.5" X 0.5" intervals respectively at the Equator. For users requiring data spaced in meters on the ground, the UTM-Standard format is used where gridded data conform to the UTM coordinate system. DEMs may also be produced in local table coordinates for digital terrain modeling. Modified UTM or Geographic formats may be produced to satisfy a particular users needs. In all cases, the size of the data set and resolution of the data are determined by the primary users requirements.

BACKGROUND

In the late 1960's, a primary product developed at HTC was the plastic relief map. These were produced by drawing a heated plastic map down over a hand carved model of the earth's surface in a vacuum frame, thus forming the plastic relief map. Producing the carved model of the earth's surface was a very time consuming and costly process involving many labor intensive tasks. To improve the process of generating these terrain models, DMA initiated R&D efforts aimed at digital collection of terrain information which, in turn, would support automatic production of terrain models.

In 1968, DMA procured the Digital Graphics Recorder (DGR) for digitizing contours from existing map manuscripts, and developed software for interpolating DEMs from the contour data. In the same timeframe, the Universal Automatic Map Compilation Equipment (UNAMACE) was developed to automatically generate DEMs from stereo photography. The original intent of these DEMs was for driving a milling machine for producing terrain models. However, the desirability of the DEM soon overshadowed the plastic relief map as a major product of the Defense Mapping Agency. As digital processing capabilities increased, so increased the demand for greater quantity, accuracy and diversity of DEM products. DEMs became highly desirable for such applications as: quantity determinators for engineering, slope analysis for cross country movement, contour, perspective scene and shaded relief generation for mapping, lines of site analysis for communications, scene generation for simulators, and guidance systems for advanced weapons. As requirements for DEMs increased, the production capabilities at DMA to meet these requirements, also increased.

PRESENT PRODUCTION SYSTEM

DATA COLLECTION

Data is collected at DMA on one of three collection systems; the Automatic Graphic Digitizing System (AGDS), the Upgraded UNAMACE, and the Pooled Analytical Stereo System (PASS).

AGDS. The AGDS is a fully integrated manuscript digitizing system consisting of a flatbed laser scanner that collects contours, drains, ridges, lakes and other supplemental information in raster form at a resolution of D.001". After scanning, the data is vectorized, then edited and tagged in vector form prior to post processing into DEM formats.

Upgraded UNAMACE. The Upgraded UNAMACE is a photogrammetric collection instrument that automatically correlates stereo photography and collects data in a uniform matrix in a local orthogonal coordinate system, with minimum operator intervention. Limited editing may be performed on the Upgraded UNAMACE prior to post processing.

PASS. The PASS is another photogrammetric collection instrument used to collect data in a more interactive environment. Data collected with the PASS consists of Non-Uniform Matrices in profile form, geomorphic data (drains and ridges), lakes and double line drains, and control points. Like the UNAMACE, data may undergo limited editing prior to post processing into DEM formats.

POST PROCESSING

Post processing of DTED refers to all processing of data between collection and output as verified user specified products. Post processing occurs as a series of independent batch processes on 2 UNIVAC 1100/81 mainframe computers. These processes include the contour to grid interpolation functions for AGDS vector data, coordinate transformation and interpolation functions to transform data between local table, geographic, local orthogonal and UTM coordinate systems, and data integration functions to produce standard DMA products or special user formats. Editing functions are also performed in batch mode to remove anomolies and discrepancies in the data, perform hole fill operations, detect and remove noise in the data, and perform paneling operations to insure a smooth continuous surface between data sets. For verification throughout the post processing phase, a variety of plot files are generated for offline plotting on raster and vector plotting systems; these include profile plots, convolution or edge enhanced plots, contour plots, grey scale and shaded relief plots. Control point analyses are also performed to insure the accuracy of the final product.

VALIDATION AND QUALITY CONTROL

To insure the quality of the data produced at HTC, the Image Manipulation Station (IMS) and the Sensor Image Simulator (SIS) are employed. These systems provide the capability of generating advanced image displays for viewing DEMs prior to insertion into the DMA data bank or delivery to digital data users.

DEFICIENCIES IN THE CURRENT SYSTEM

DMA has in place the collection systems, and validation and quality control systems capable of meeting future DTED production requirements. However, deficiencies exist in post processing that must be addressed to satisfy those requirements. Problems (such as the response time associated with batch processing which often requires a 24 to 48 hour waiting period to review the results of an edit process, only to find that a keypunch or tape error aborted the run) severely affect throughput requirements. Limitations associated with the current plotting capability also affect throughput. When errors not detected on existing plots appear in a quality control system, the data set must be returned to post processing for correction; thus disrupting production. There are other deficiencies associated with production flow, file handling, editing procedures and processing algorithms which cause post processing in a batch environment to severely impact production.

TES/EMPS - THE FUTURE IN DTED PROCESSING

The previous sections have described the evolution of both the requirements for DTED and the current methods to produce it. As with any evolutionary process, piecemeal requirements were met with piecemeal solutions that were minimum cost modifications of existing capabilities. Often these solutions were mandated to utilize existing equipment. Experience has shown that such constrained Darwinian systems may minimize incremental costs, but generally at the expense of both total system capabilities/effectiveness and operation costs. Such production systems seem to agonizingly crawl to untenable positions, burdened by untractable loads, achieving only expensive mediocrity.

The TES/EMPS procurement at once revealed the potential benefits of a systems approach to expanding requirements beyond all expectations (anticipated, and yet unspecified). The subheadings of this section document the chronological steps that have been followed to develop a DTED system that not only meets today's processing requirements but is also guaranteed to cost-effectively expand in the future. The first subsection summarizes the system processing requirements. The second subsection covers design issues and objectives that are explicitly detailed and used to shape the design philosophy presented in the third subsection. The hardware design based upon this derived philosophy is covered in the fourth subsection. The final subsection completes the development cycle, summarizing the final systems operations characteristics.

System Requirements: Goals to Grow On

The TES/EMPS will perform all processing of digital elevation data from photographic and cartographic sources - from collection through output as verified digital elevation products. The TES/EMPS will address for the first time in a single integrated system the seven primary functions required in production processing of DEM data:

(1) Edit anomalies and discrepancies
(2) Convert contour data to matrix data
(3) Transform and interpolate DEMs to other coordinate systems
(4) Merge DEM's into product formats
(5) Edit and verify product DEMs
(6) Perform File and Job Management functions
(7) Perform rigorous quality control.

The TES/EMPS system was specified to be an integrated system composed of off-the-shelf, commercially available computers and peripherals of modern design, proven system software, appropriate interactive graphics software, and specific applications software. The unique requirement with potentially the largest production payoff is for an integrated job (or DEM project) management approach that optimizes equipment utilization to achieve the throughput rates mandated by the DMAHTC data user community.

Design Objectives: Guidance From the Past

Analysis of the requirements in light of previous production system acquisition experience leads to design objectives. The purpose of developing explicit design objectives, separate and distinct from design requirements, is to maximize system applicability; not only at time of delivery, but also for the entire working life of the system. Further, the disappointments and frustrations encountered in previous developments can be minimized, if not avoided.

The processing of digital elevation data to address all seven of the functions listed in the previous section requires the marriage of an extremely capable background processing system with an interactive processing system linked to a common data base and function set.

The TES/EMPS requirements are further constrained by four additional "production plant" considerations: reliability, maintainability, implementation risk, and expansion potential. Reliability, risk reduction and growth are the driving forces behind the DMAHTC requirement for use of off-the-shelf, commercially available components of modern design. Maintainability is achieved in three ways: good design, experienced maintenance personnel, and availability of spare parts.

The risk level for the design and implementation methodology of a project with the scope of TES/EMPS is a critical issue. The magnitude of cost, production workload, and system effectiveness demands that a low risk approach be selected.

The observed growth in requirements for DTED, from both increased demands resulting from current applications and new applications (anticipated and unexpected), strongly points to TES/EMPS expansion. The expansion capability must address not only increased throughput, but also increasing sophistication of processing.

In summary, the key issues and objectives of the TES/EMPS design are:

- o Extensive background processing capability
- o State-of-the-art interactive graphics
- o Integrated data bases and processing functions through a common file management strategy
- o Sophisticated job management for optimal resource utilization
- o High performance
- o High system reliability
- o Continuing system maintainability
- o Low-risk implementation
- o Extensive, low-cost, and long-term expansion capability

Design Philosophy: A Roadmap to Success

The system requirements and design objectives form a firm foundation upon which to build the high throughput production system desired.

Intergraph Corporation is one of the leading manufacturers of Computer Aided Design and Computer Aided Manufacture (CAD/CAM) Systems in the world. For interactive mapping applications, Intergraph is the leader.

A majority of the capabilities required for TES/EMPS are currently available in Intergraph systems. The TES/EMPS hardware requirements match closely with Intergraph hardware currently in production and use. New items brought into production by Intergraph in 1983, in combination with previous production equipment, will satisfy all of DMA's hardware requirements. Many software requirements are met by Intergraph's base level graphics and data base facilities contained in the Interactive Graphics Design Software (IGDS) and Data Management and Retrieval System (DMRS). Many unique DMA DTED processing functions are also contained in existing DMA production programs and other vendor software.

With this background, the lowest risk, lowest cost design philosophy, or technical approach to achieve a fully responsive, high performance system is built around the following critical design features:

- o Maximum use of off-the-shelf hardware and software concepts
- o Use of off-the-shelf government and industry application software and algorithms
- o Integrated systems design that minimizes special or new software, algorithms or data formats

These critical design features will insure the final system has:

- o maximum reliability
- o assured maintainability
- o lowest equipment cost through minimum R&D
- o easy integration of advances in the commercial sector
- o highest assurance of system growth potential
- o minimized dependence upon unproven algorithms

Hardware Design: A Basis for Achievement

Analysis of the DMA DTED processing requirements show that they logically fall into three major categories: Interactive, Batch and Production Control Processing. The interactive processing includes quality control, data set merging, paneling and error correction. All processes, experience has shown, require a man in the loop for successful performance. The batch processing requirements include input/output operations, coordinate transformations, format conversions (such as, contour-to-grid), proof plot preparation, anomaly/discrepancy detection and other automated quality control aids. In general, these are fully automated CPU intensive background processes. The production control processing contains both fully automated tracking and scheduling operations, as well as interactive capabilities for system work assignments, status reporting and work flow control.

The basic hardware components of the TES/EMPS are the VAX 11/780, the workstation, the File Processor, the Graphics Processor, the Banded Vector to Raster Converter (BVRC) and Internet. These major components are supplemented with three billion bytes of mass storage and a full complement of the usual peripherals; line printers, tape drives, plotters, etc. The TES/EMPS will use three of Digital Equipment Corporation's 32 bit Vax 11/78D CPU's. Each of fourteen TES/EMPS workstations will be equipped with one 19" color display and one 19" black and white display, both with a resolution of 1280 by 1024 picture elements. The color display will be supported by 16 planes of refresh memory and will selectively display 256 grey scales or 256 levels of intensity for each color; red, green and blue. The Graphics Processor has been added to assume the background computational load. This 64 bit, programmable array processor, as a pipeline unit, is designed to operate in parallel with the central processor. The BVRC performs a high-speed vector to raster conversion enabling the plotter to operate at maximum speed. The File Processor is an intelligent disk controller which scans a disk at rotational speeds, analyzing and transferring to the Central Processor only the data that meet desired search criteria. This search function represents a second critical burden removed from the Central Processor, to enable the overall system to match the users' design pace. Internet is a local area CPU network for intersystem file access and transfer.

Two Intergraph 780 Data Processing Systems, connected to 12 Intergraph graphics workstations, will form the hardware backbone of the terrain editing system. Each Data Processing System will be equipped with two graphics processors. A third Intergraph system, with additional workstations, will be dedicated to control functions within TES/EMPS. This system will manage the project queue, parceling out jobs to the other systems, and will manage operations information and perform a large part of the matrix processing to support terrain editing and analysis. The third system will perform batch matrix processing and plot preparation work with the assistance of another graphics processor and two BVRCs, respectively.

Operational Characteristics: Realization of a Goal:

The following example scenario provides insight into the flexibility and power available for total DTED production processing with implementation of the TES/EMPS design.

Data enter the system via magnetic tape and are placed in mass storage by File Management. Based upon Production Manager Job Supervisor inputs, the Job Management subsystem creates a job step sequence and schedule.

Based upon the job step sequence, the Job Management subsystem directs the pre-edit software to acquire the data files via File Management and processes the data. The pre-edit software produces any required plots. Any background matrix processing is completed on an available VAX, as required for a given task. Upon completion, the data are transferred over the interprocessor communications link to the File Management subsystem on one of the edit system VAX's . Job Management controls this transfer and selects which edit VAX is to receive the data based upon resource availability. The data are entered into the edit VAX's mass storage by the edit VAX's File Management subsystem. Once the assigned editor selects the edit job step for these data from his work queue, the DEM Edit applications software uses File Management to access the data. Based upon editor commands, the DEM edit software establishes the type of data displays (either vector or image) most useful for the assigned edit functions. The transmission rates for the image format edit displays are assured through utilization of Intergraph's parallel I/O interface to the edit station. Real time three-dimensional rotation, zoom and pan of the terrain presentations (from 2 to 8 views) are available so that editors will be able to effect changes dynamically and see the results immediately. Project continuity will be vastly improved, with dramatic increases in accuracy and productivity. The decrease in feedback time alone is expected to provide a strong boost to employee job satisfaction. Upon completion of edit, the corrected data files are transferred back to VAX 1 via the interprocessor communication link. As appropriate, either post or matrix processing (for coordinate transformation) is scheduled to complete processing of the active files. At completion of this and any other edit job steps specified by the Production Supervisor and scheduled by Job Management, the data are stored on VAX 1 mass storage by File Management. Job Management schedules and accomplishes the generation of plot tapes, quality control sessions at the Production Supervisor's station, output via the I/O processor of archive tapes, and finished DMA product tapes along with job schedules and status reports.

CARTOGRAPHIE ASSISTÉE PAR ORDINATEUR POUR LE RECENSEMENT DU CANADA

D. Ross Bradley
Statistique Canada
Division de la géographie
Édifice Jean-Talon
3e étage, Section B-7
Ottawa, Ontario
K1A 0T6

RÉSUMÉ

Le recensement quinquennal du Canada est une énorme entreprise. Pour l'effectuer de façon ordonnée, il faut équiper chacun des 35,000 à 40,000 recenseurs d'une carte qui assure une livraison exacte et complète des questionnaires. Lorsque le moment vient de publier les résultats du recensement, qui sont diffusés à divers niveaux de ventilation géographique, il faut des cartes de référence pour situer les limites des zones visées. Les résultats sont aussi diffusés sous forme de cartes thématiques qui suppléent en quelque sorte à la présentation classique sous forme de tableaux. L'informatisation des cartes de collecte, des cartes de référence et des cartes thématiques du recensement a atteint un stade avancé à Statistique Canada.

INTRODUCTION

Le recensement du Canada, que l'on dit souvent être la plus grande activité en temps de paix, effectué par le gouvernement du Canada, a lieu une fois tous les cinq ans. Il vise à dénombrer toutes les personnes demeurant au Canada à leur domicile habituel le jour du recensement. Pour mener à bien une tâche de cette envergure, quelque 35,000 à 40,000 recenseurs doivent livrer des questionnaires à environ 9,300,000 ménages au cours d'une période très courte. De plus, pour assurer le recensement de toutes les régions, il est nécessaire de remettre des cartes aux recenseurs afin que toutes les régions ne soient visitées qu'une fois et une fois seulement. De là le besoin de cartes de collecte. Le recensement doit aussi publier les résultats de ses trouvailles. Les données démographiques paraissent au niveau de nombreuses régions géographiques et des cartes de référence existent pour toutes ces régions. Les cartes thématiques constituent une autre façon de présenter les données. Pour le recensement de 1981, ces cartes se trouvent dans la Série d'atlas métropolitains (pour 12 grandes régions métropolitaines) qui viendra compléter les bulletins sur les profils des secteurs de recensement.

Le présent document décrit la cartographie assistée par ordinateur pour ces trois genres de cartes.

CARTES DE COLLECTE

Il est évident que les recenseurs ont besoin d'un volume impressionnant de cartes à grande échelle. Étant donné que les recenseurs ne sont habituellement pas formés à lire des cartes, toutes celles qu'ils utiliseront devront être très claires, très précises et faciles à lire. La production de cartes pour la collecte des données au recensement coûte très cher. Lors du recensement de 1981, le programme de cartographie a nécessité quelque 80 années-personnes en commis et dessinateurs et environ $1,500,000. Juste avant le recensement de 1981, on a procédé à un examen minutieux de ce programme afin de pouvoir économiser des ressources humaines et financières en faisant appel à une technologie informatique et en intégrant les activités. Ainsi donc, commença la création de cartes par des moyens informatisés, aux fins de la collecte.

Pour que les cartes soient produites par des moyens informatisés, il est évident qu'il faut disposer d'un fichier ordinolingue. Il existe un fichier à cette fin que l'on appelle le fichier principal de la région (FPR), qui est composé d'un réseau de rues en numérique et de renseignements connexes comme le nom des rues, les tranches d'adresses et d'autres caractéristiques géographiques. Il y a un FPR pour presque tous les noyaux urbanisés de chaque centre de 50,000 habitants ou plus, et le Canada en compte 37; mentionnons entre autres, des endroits comme North Bay qui compte un peu plus de 50,000 habitants et des grandes villes comme Toronto et Montréal dont la population du noyau urbanisé atteint globalement 2,500,000 habitants. Les FPR renferment une représentation ordinolingue logique de toutes les rues des villes et d'autres caractéristiques choisies comme les voies ferrées, les cours d'eau, les limites de municipalités et bien d'autres. Les FPR donnent une référence géographique des coordonnées de chaque rue, tranche d'adresses, côté d'îlot et centroïde dans la région en question. Une carte tracée par ordinateur constitue un produit important du FPR. À l'aide du programme utilitaire MAPMAKR, il est possible de dresser des cartes indiquant les réseaux de rues sous forme de lignes simples (voir la figure 1).

Étant donné que la collecte des données et la couverture complète du territoire représentent deux des plus importantes activités du recensement, la carte à lignes simples dressée à partir du FPR n'est pas assez précise à cette fin. Aussi a-t-on créé un programme visant à produire une carte des rues sous forme de lignes doubles à partir du réseau à lignes simples (voir la figure 2). Un système expérimental a été mis au point et quelque 200 cartes ont été remises aux recenseurs pour le recensement de 1981. Une évaluation ultérieure de ces cartes a permis d'établir qu'elles suffisaient amplement à la tâche. Le coût des cartes produites par des moyens informatiques pour le prototype était assez proche de celui des cartes que l'on dressait par des méthodes traditionnelles; cependant, la main-d'oeuvre nécessaire a été beaucoup moins imposante et il a été possible de réduire encore plus les coûts. Au cours de la dernière année, Statistique Canada a mis au point un système de production visant à dresser environ 8,000 cartes de secteurs de dénombrement pour le prochain recensement. On en évalue les coûts à l'heure actuelle à environ le tiers de ce qu'ils étaient pour le prototype en 1981. Parmi les autres avantages, mentionnons une cohérence accrue au niveau de la collecte, du traitement et de la documentation des cartes de base.

Bien que ces cartes se répercutent directement sur les activités du recensement, d'autres enquêtes à Statistique Canada, comme l'enquête sur la population active, les trouveront utiles tout comme d'ailleurs le ministère des Postes, et le bureau du Directeur général des élections et ses homologues provinciaux.

Soit dit en passant, il est intéressant de noter que les fichiers principaux de régions n'ont pas été créés initialement à des fins de cartographie. Le système de géocodage, dont les FPR sont la base, a été mis au point au début des années 70 afin d'assurer une grande souplesse à l'extraction des données du recensement attribuées à ces centroïdes de côté d'îlot pour des régions spécifiées par les utilisateurs (voir figure 1). Les cartes de collecte du recensement sont toutes dérivées de ce système et pourront très bien devenir l'objet principal des FPR.

CARTES DE RÉFÉRENCE

Aux fins du recensement, une carte de référence est une carte qui permet d'identifier et de trouver les limites d'une région géographique pour laquelle des données de recensement existent. Des cartes de référence sont publiées pour les divisions de recensement (comtés, municipalités régionales, districts régionaux, etc.), les subdivisions de recensement (villes, villages, townships, cantons, réserves indiennes, etc.), les régions métropolitaines de recensement, les agglomérations de recensement, les secteurs de recensement et les circonscriptions électorales fédérales. Des cartes de référence comme celles qui sont présentées aux figures 3, 4 et 5 ont été produites de façon traditionnelle depuis déjà de nombreux recensements.

Dans le cadre de son programme de publications du recensement de 1981, Statistique Canada a produit pour la première fois une série de cartes de référence à l'aide d'un ordinateur. Le bulletin intitulé Cartes de référence: Divisions et subdivisions de recensement (n^{o} 99-907 au catalogue) du recensement de 1981 renferme quelque 41 cartes produites à l'aide d'un ordinateur. La production de ces cartes a fait appel à des techniques informatisées et manuelles de cartographie en intégrant trois fichiers en numérique.

a) Le fichier des limites de subdivisions de recensement (municipalités) obtenues en convertissant en numérique les frontières tirées des cartes de base à une échelle de 1/50,000^{e} et 1/250,000^{e} du ministère de l'Énergie, Mines et Ressources Canada.

b) Le fichier des noms et des codes de la Classification géographique type (CGT) de Statistique Canada.

   Nota: Les fichiers a) et b) étaient des composantes du fichier géographique principal du recensement de 1981 (FGPR).

   Le FGPR comprend tous les renseignements nécessaires à l'extraction des données du recensement pour toute région géographique normalisée au Canada. Grâce à une capacité de contrôle géographique, ce fichier peut vérifier le code de chaque secteur de dénombrement dans la région géostatistique qui lui est propre.

c) Le fichier des cours d'eau et des rives convertis en numérique à partir des cartes de base à une échelle de 1/2,000,000 du ministère de l'Énergie, Mines et Ressources Canada.

Le fichier des limites de la projection transverse de Mercator (PTM) (voir a) ci-dessus) a été converti aux coordonnées du système conforme de Lambert puis superposé au fichier des cours d'eau; on a ensuite éliminé directement sur l'écran à tube cathodique toute ligne inutile ou dédoublée. Ainsi, si une limite de municipalité ressortait dans un cours d'eau, on l'effaçait de façon à n'indiquer que la rive. Les noms provenant du fichier de subdivisions de recensement ont été automatiquement placés au centroïde de chaque municipalité, permettant ainsi de relier le code géographique des deux fichiers. S'il y avait trop de noms, ils étaient replacés interactivement afin de produire un espacement satisfaisant. Les cartes du Yukon, des Territoires du Nord-Ouest, de l'Île-du-Prince-Édouard, de la Nouvelle-Écosse et du Nouveau-Brunswick ont été produites à l'aide de cette méthode (voir la figure 6). Les autres cartes de cette série ont été dressées en faisant appel à un plus grand nombre de commis et en appliquant des opérations cartographiques plus conventionnelles. On a utilisé le fichier des limites des municipalités, des cours d'eau et des rives de pair avec les données de l'ordinateur. Les lignes inutiles étaient éliminées à la main plutôt qu'en direct alors que les noms et les codes étaient placés à la main à l'aide d'une liste et d'un croquis informatisés sur lesquels les noms figuraient au centroïde de chaque municipalité (voir la figure 7).

Toutes ces cartes ont été produites en trois couleurs: rouge pour les limites des 266 divisions de recensement et leurs codes géographiques normalisés; bleu pour les cours d'eau et les rives; et noir pour les limites et les noms des 5,710 subdivisions de recensement (et leurs codes pour plus de 2,600 de ces dernières si la subdivision de recensement est une subdivision de recensement unifiée) ayant servi à la diffusion des données du recensement de 1981.

CARTES THÉMATIQUES

Bien que la plupart des données du recensement paraissent sous forme de tableaux et soient publiées dans les bulletins, la technique d'affichage des données statistiques à l'aide de cartes thématiques est très efficace. À la suite d'un prototype d'atlas sur la région métropolitaine de recensement d'Ottawa-Hull, produit à la suite du recensement de 1976, on a décidé de produire une Série d'atlas métropolitains pour le recensement de 1981 qui décrirait, sous forme de cartes thématiques et de graphiques, la répartition des diverses caractéristiques de la population pour les régions métropolitaines de recensement de Toronto, Montréal, Vancouver, Edmonton, Calgary, Regina, Winnipeg, Hamilton, Ottawa-Hull, Québec, Halifax et St. John's.

Les atlas de cartes thématiques servent beaucoup en planification urbaine, en service social et en éducation, surtout pour se faire une idée générale des données et de leur répartition géographique qui peut alors être suivie avec beaucoup plus de détails. Les atlas métropolitains du recensement de 1981 renfermeront un grand nombre de cartes et de graphiques qui illustreront une diversité de thèmes ou de variables du recensement (voir les figures 8 et 9) pour une région métropolitaine donnée. La cartographie se fait au niveau du secteur de recensement (c'est-à-dire une petite unité géostatistique comptant en moyenne 4,000 habitants et qui est établie dans les communautés urbaines de 50,000 habitants et plus) et elle sert à compléter les bulletins des profils des secteurs de recensement. La cartographie permet de révéler les tendances spéciales qui ne sont pas immédiatement apparentes dans une région métropolitaine lorsqu'on les présente sous forme de tableaux.

La Série d'atlas métropolitains est un produit à la fine pointe de la technologie actuelle. Des progrès au niveau des procédures d'analyse des données, des logiciels de cartographie informatisée et de matériel de tracés ont permis de produire ces atlas. Les programmes destinés à la cartographie informatique ont été fournis par le GIMMS (Système de cartographie et de manipulation de l'information géographique), tandis que l'analyse des données des secteurs de recensement fait appel à une série de programmes informatiques qui aident à choisir les intervalles de classe appropriés et la densité des points sur les cartes et les graphiques. Chaque page est entièrement produite sur un traceur à tambour à haute vitesse GERBER, modèle 4442, qui utilise une tête spéciale par laquelle un faisceau lumineux concentré traverse une pellicule photographique enroulée sur le tambour. Comme les pages sont imprimées à leur grandeur réelle, les pellicules peuvent être envoyées directement au service d'impression des bulletins, évitant ainsi toute perte de qualité ou étape intermédiaire. Les cartes se présentent sous forme monochromatique de façon à faciliter la reproduction et à réduire les coûts d'impression.

La cartographie assistée par ordinateur à Statistique Canada est le fruit d'un travail conjoint de la Sous-division de la géocartographie, Division du traitement des données, et de la Division de la géographie. Le matériel, le logiciel et les services de cartographie des programmes décrits ont été fournis par la Sous-division de la géocartographie. Cette dernière compte d'ailleurs les spécialistes sur place pour répondre aux besoins du recensement et des autres secteurs de Statistique Canada.

En guise de conclusion, disons simplement que la cartographie assistée par ordinateur a fait des progrès énormes pour le recensement à Statistique Canada au cours des dernières années. L'uniformité des produits et les économies qui peuvent être réalisées grâce à l'automatisation dépassent de loin toutes les lacunes esthétiques que l'on peut retrouver dans ce genre de cartes. Il ne fait aucun doute qu'il nous reste encore beaucoup de chemin à parcourir dans ce domaine. Les cartographes devraient relever le défi et montrer plus d'initiative dans la cartographie informatisée. Grâce aux techniques qui leur sont particulières à chacun, le cartographe et l'informaticien peuvent former une équipe des plus enviables.

BIBLIOGRAPHIE

Bradley, D.R. 1981, _Cartes de collecte assistées par ordinateur pour le recensement du Canada de 1981_, pp. 1-6

Statistique Canada, 1972, _GRDSR: Renseignements par petits secteurs_, pp. 1-23

Puderer, H.A. 1982, Census Subdivision Reference Maps: _Task Documentation_

Ross, G. 1983, _Guide de préparation des atlas métropolitains_, pp. 1-25

Sous-division de la géocartographie, 1982, _Hardware/Software Descriptions_, pp. 1-11

Wellar, B.S. 1982, _Towards a Strategic/Master Plan for Coordinating Geographic, Cartographic and Graphic Census Activities of the Geography Staff, Statistics Canada_, pp. 1-46

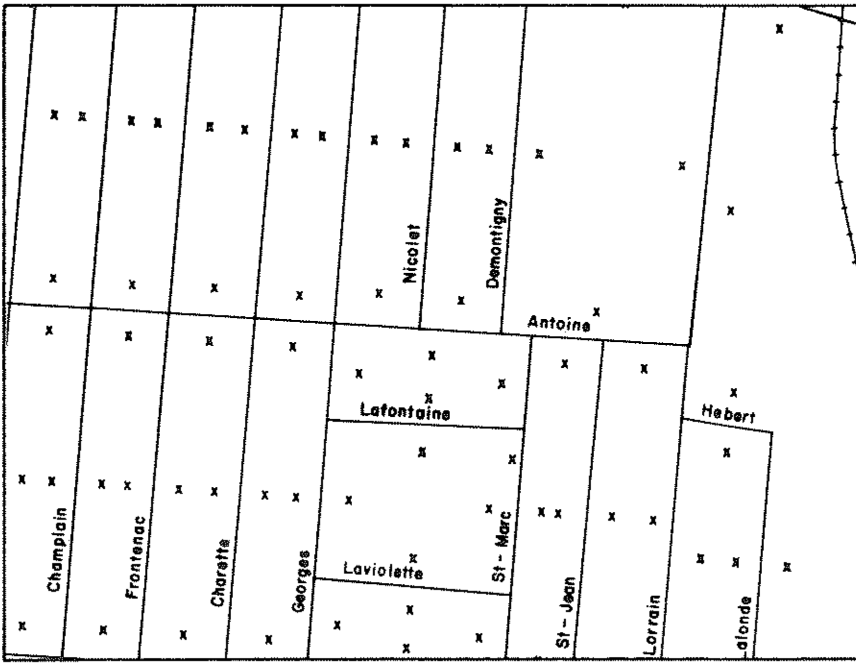


Figure 1. FICHIER PRINCIPAL DE LA RÉGION IMPRIMÉ (LIGNE SIMPLE)

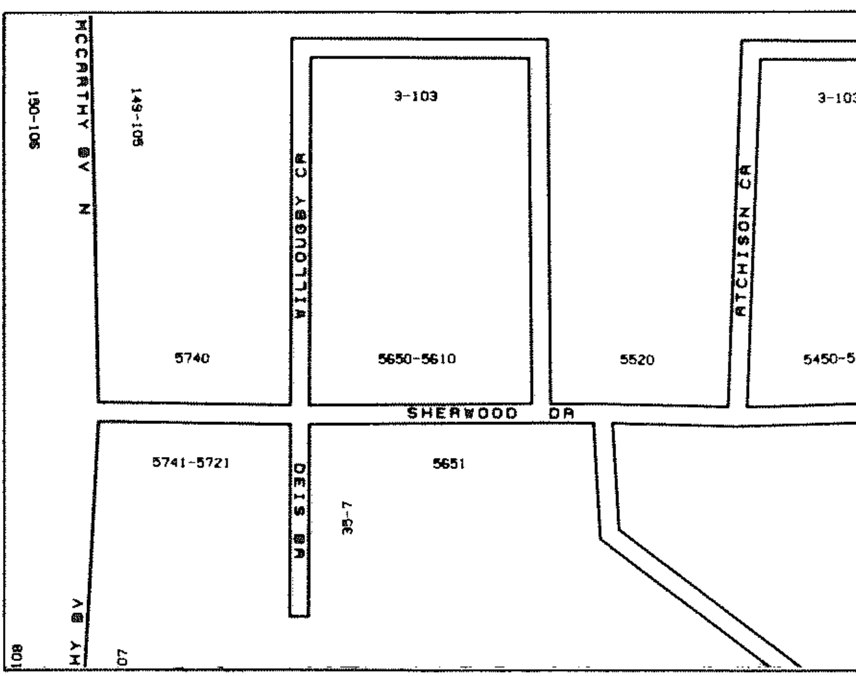


Figure 2. CARTE DE COLLECTE DESSINÉE PAR ORDINATEUR

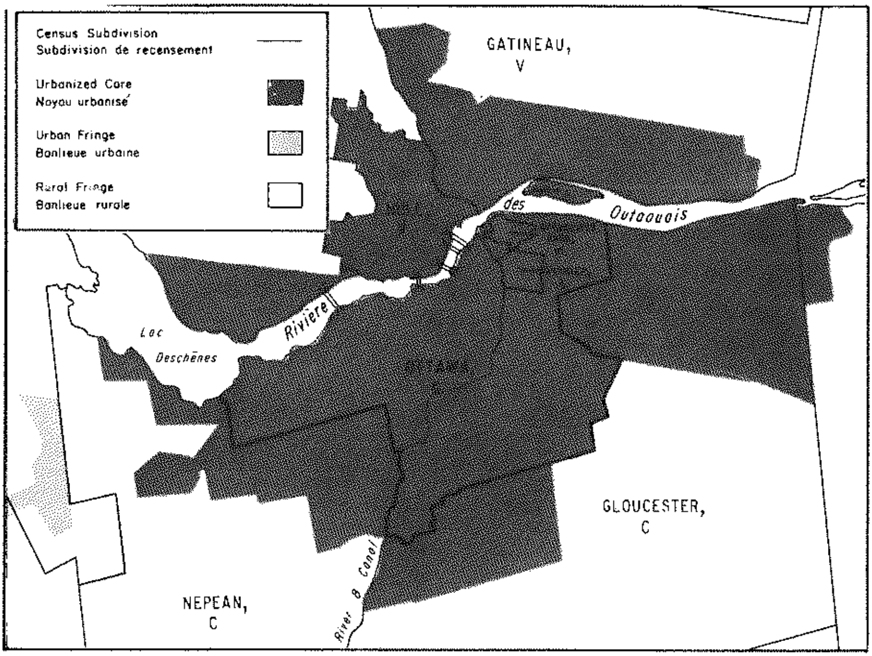


Figure 3. RÉGION MÉTROPOLITAINE DE RECENSEMENT

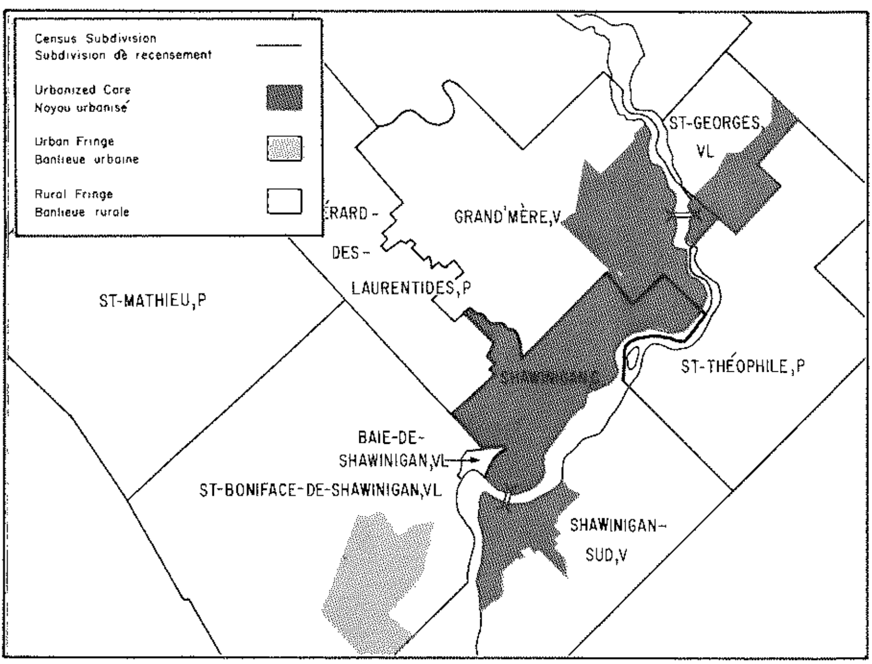


Figure 4. AGGLOMÉRATION DE RECENSEMENT

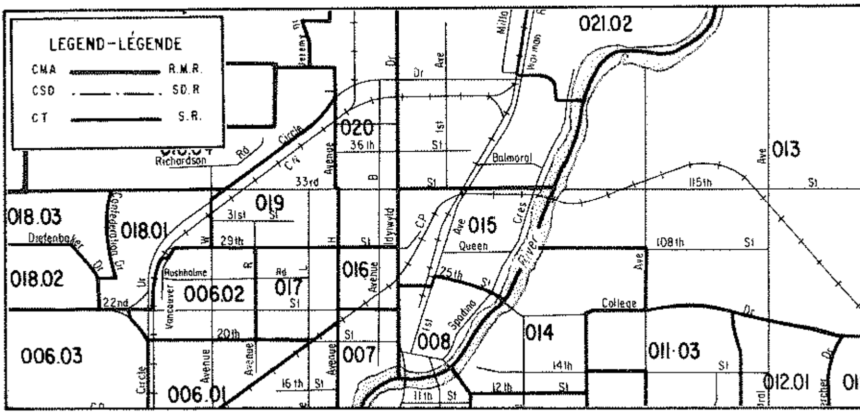


Figure 5. CARTE INDEXE DES SECTEURS DE RECENSEMENT

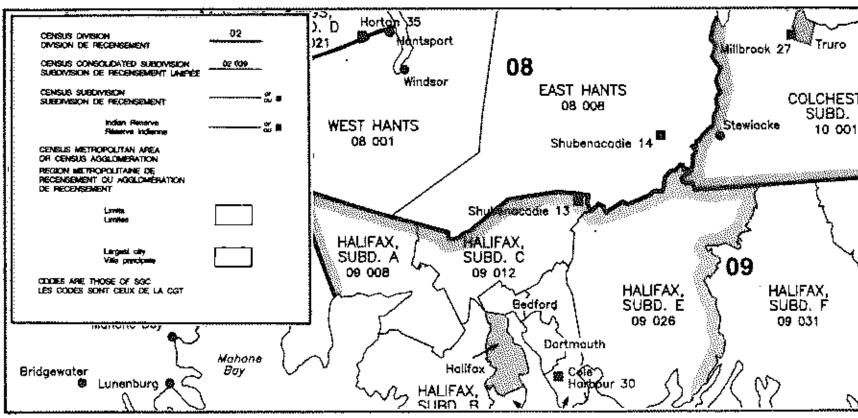


Figure 6. SUBDIVISION DE RECENSEMENT

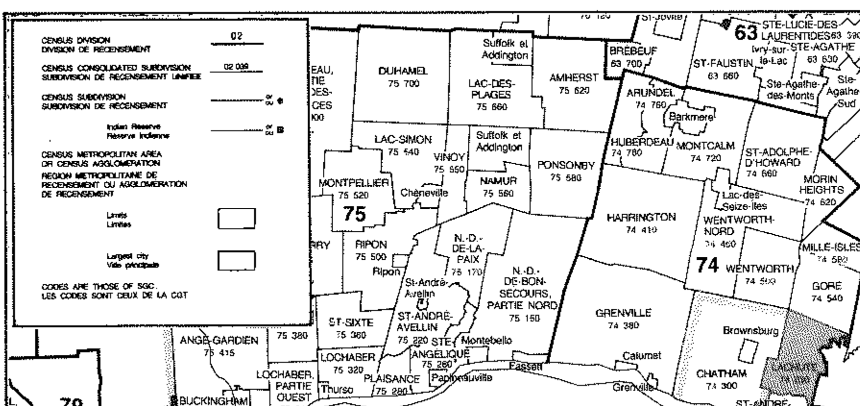


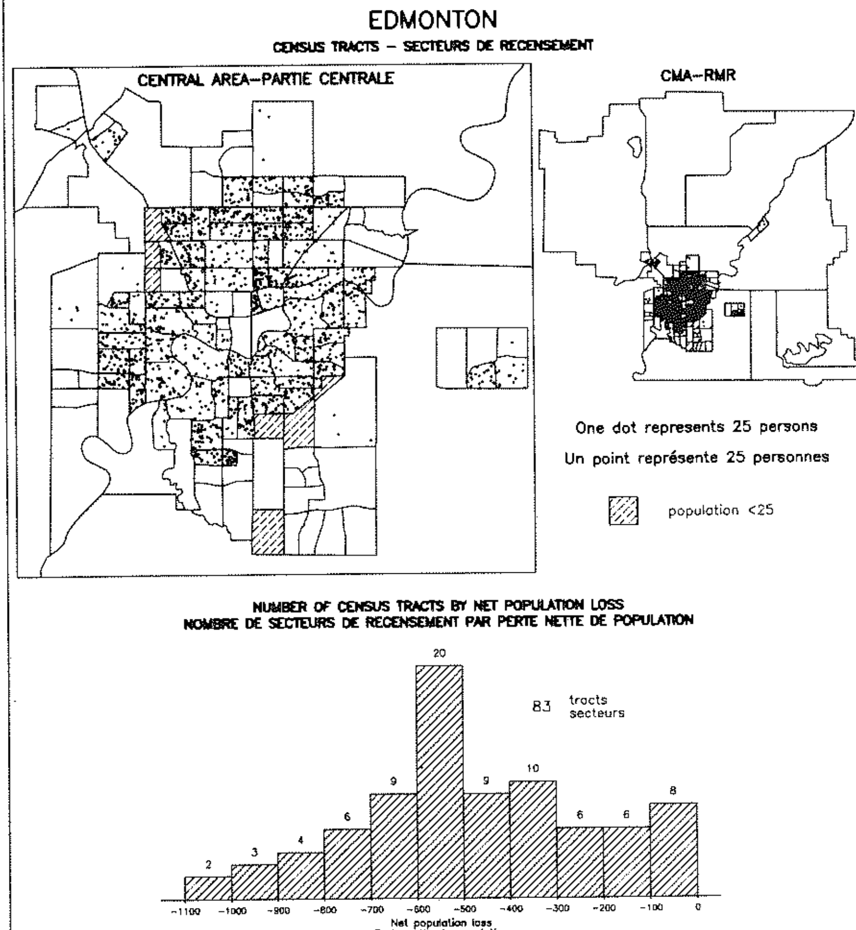
Figure 7. SUBDIVISION DE RECENSEMENT

NET POPULATION LOSS, 1976 – 1981

Census tracts which experienced a net loss of population between 1976 and 1981 are highlighted. Net losses are shown in absolute numbers of persons.

PERTE NETTE DE POPULATION, 1976 – 1981

Les secteurs de recensement subissant une perte nette de population entre 1976 et 1981 sont mis en évidence. Les pertes nettes sont données en nombres réels de personnes.



SOURCE: 1981 CENSUS OF CANADA
PRODUCED BY STATISTICS CANADA, 1982.

SOURCE: RECENSEMENT DU CANADA DE 1981
ÉTABLIE PAR STATISTIQUE CANADA, 1982.

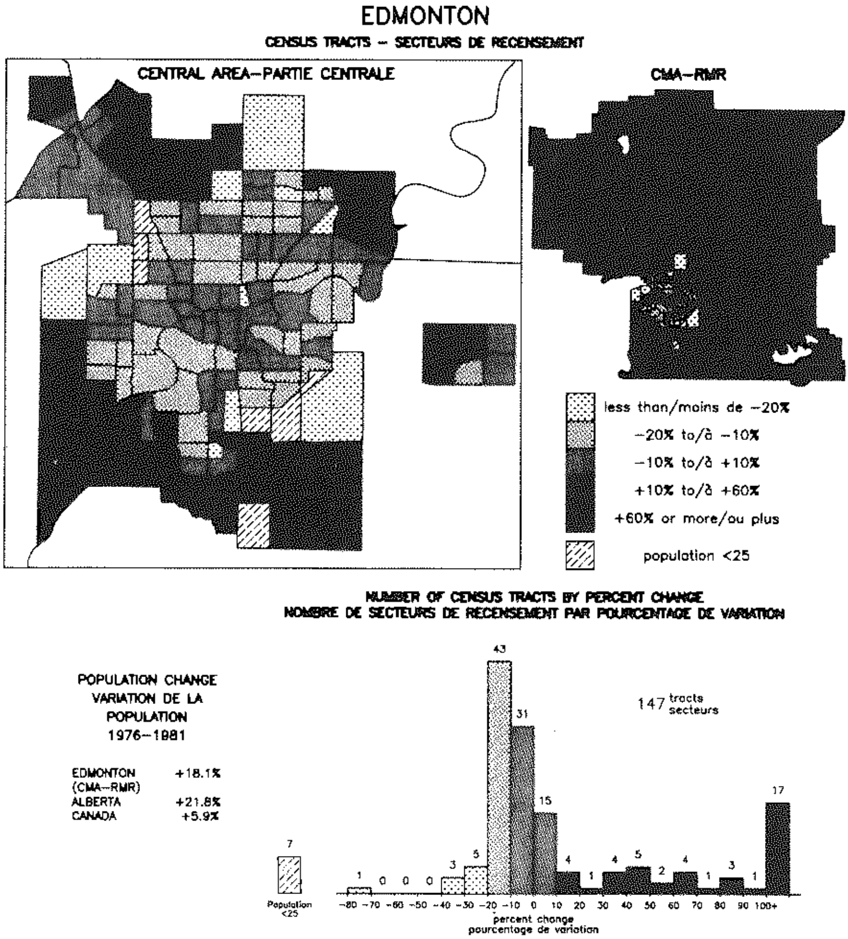
Figure 8.

POPULATION CHANGE, 1976 – 1981

VARIATION DE LA POPULATION, 1976 – 1981

Population change between 1976 and 1981 is expressed as a percentage of the total 1976 population for each census tract in the census metropolitan area (CMA). Comparative figures are provided for the CMA, the province and Canada.

La variation de la population entre 1976 et 1981 est exprimée en pourcentage de la population totale de 1976 de chaque secteur de recensement de la région métro-politaine de recensement (RMR). Les données comparatives sont produites pour la RMR, la province et le Canada.



SOURCE: 1981 CENSUS OF CANADA
PRODUCED BY STATISTICS CANADA, 1982.

SOURCE: RECENSEMENT DU CANADA DE 1981
ÉTABLIE PAR STATISTIQUE CANADA, 1982.

Figure 9.

NOTA

Les figures 3, 4, 5, 6, 7 sont disponibles en couleur et peuvent être obtenues de l'auteur.

HYDROGRAPHIC SURVEY REQUIREMENTS SYSTEM (HYSUR)

Chung Hye Read
Defense Mapping Agency
Hydrographic/Topographic Center
Washington, D.C. 20315

ABSTRACT

The Defense Mapping Agency (DMA) has no automated system with which to determine the survey precedence of the 8.2 million square nautical miles of worldwide coastal areas which require hydrographic surveying. In response to this vast requirement, the DMA Hydrographic/Topographic Center (HTC) is developing the Hydrographic Survey Requirements System (HYSUR). This software system will enable managers to make timely, systematic decisions, thereby enhancing the utility of the Naval Oceanographic Office (NAVOCEANO) survey vessels. This report discusses the practical applications of this system which will:

a. Identify hydrographic data deficiency areas.

b. Establish a rank order for required survey areas considering economic, political, military, and physical factors.

c. Provide an alternative planning capability to answer management-level "What if" questions, such as: "What if a given route is blocked?", or "What if merchant traffic is disregarded?".

d. Produce coastal survey requirements based on the results of the ranking procedures, available resources, and estimated cost.

INTRODUCTION

The Hydrographic Survey Requirements System (HYSUR) is a planned software system that will enhance the Defense Mapping Agency's (DMA) utilization of limited hydrographic survey resources to accomplish needed surveys. HYSUR will provide an automated method to assess existing source material coverage of the required areas, and to establish weighting factors based on economic, political, geomorphological, and military considerations in order to determine hydrographic survey priorities for required geographic areas.

An automated file will be maintained on both the mission areas, which will be defined, and their subdivisions, called cell areas. This will be a static data file that identifies cell area size, as well as the degree of safe passage for specific coastal areas. A dynamic data file containing merchant and military traffic volume and strategic resources movement data will also be included. Additional files will include information about foreign charts, a remote sensing suitability index, and identification of countries with which the United States has mapping and charting agreements. Another HYSUR file will identify countries that the U.S. supports through the Hydrographic Survey Assistance Program. All files will be further explained in a later section.

HYSUR will be implemented on a UNIVAC 1100/62 computer utilizing the System 2000 data base management system.

BACKGROUND

The Defense Mapping Agency has a statutory responsibility* to provide accurate nautical charts of areas outside U.S. territorial waters for both civil and military maritime uses. DMA tasks the Naval Oceanographic Office (NAVOCEANO) to collect hydrographic ship survey data for specific geographic areas. Upon completion of each survey, the raw data are sent to HTC to be evaluated. The evaluated data and related information are indexed and stored for eventual use in the preparation of hydrographic products.

Prior to 1978, DMA's total coastal hydrographic data collection requirement was approximately 17 million square nautical miles. This estimate was derived by multiplying the 244,000 miles of worldwide coastline by an assumed average continental shelf width of 70 miles. To develop a more realistic estimate of coastal survey requirements, DMA conducted the Coastal Survey Requirements Study, commonly referred to as the Chubb Study, in fiscal year 1978. This study established the concept of DMA mission areas (those areas between six and 300 fathoms in depth, having heavy traffic volume and containing major coastal routes as well as significant ports and the approaches thereto). This reduced the original requirement of 17 million square nautical miles to 8.2 million square nautical miles; however, this still represented more than 200 ship-years of survey effort. The magnitude of that effort and limited survey resources still precluded a comprehensive coastal survey program within a reasonable time frame. Hence, DMA elected to develop a systematic method to establish survey priorities for hydrographic areas.

PURPOSE

Currently, when HTC is tasked to provide a hydrographic product, the areas to be charted are first plotted to determine existing DMA chart coverage of each area. Other hydrographic data in the required areas are acquired by searching various non-automated data files such as the foreign chart card file, foreign chart catalogs, and the International Hydrographic Organization (IHO) Bulletins. After evaluating all available information, a decision is made as to whether a hydrographic survey is necessary or whether existing data meet the user's requirement. If hydrographic surveys are deemed necessary, area specialists determine the ranking of the survey areas. The necessary survey planning may take up to a year.

With the increasing number of deep-draft vessels and submarines, a growing need for detailed, reliable charts exists. This need, coupled with the 200 year backlog of coastal survey requirements, results in an urgent requirement to establish rational priorities that will support an efficient program to systematically satisfy the backlog of coastal survey requirements.

---

*Title 10, U.S. Code § 2791, et seq., Chapter 167.

SYSTEM FUNCTIONS

The following system objectives have been developed in order to enhance hydrographic resource management and to assist planners in establishing priorities for survey requirements:

a. Identify deficiencies in hydrographic data by providing graphic and/or textual reports depicting shoreline and existing hydrographic data coverage for all coastal areas.

b. Establish a rank order for required survey areas using factors based on civil and military traffic volumes, commodity movements, and physical characteristics for each area.

c. Provide an alternative planning capability to answer management level "What if" decision questions, such as "What are alternative routes if a given route is inaccessible" or "What if merchant traffic is disregarded?"

d. Develop survey requirements based on the results of the ranking procedures and available survey resources.

APPROACH

Automation

Prior to the implementation of HYSUR, several data files and file indices must be automated. These files include:

a. foreign chart file, containing information on HTC's holdings of foreign charts that is currently recorded on 1500 3 X 5 inch index cards;

b. Bilateral International Agreement file, identifying countries whose charts the U.S. may reproduce; and

c. Listing of Surveys and Charts from the Hydrographic Survey Assistance Program.

In addition, the following files must be created:

a. a new related products file containing area coverage and status information on in-process and recently completed foreign charts.

b. a survey progress file containing planned, in-process, and recently completed U.S. and foreign hydrographic surveys.

c. a static data file containing mission area information, such as cell names, sizes, types, degree of safe passage and established, optimum shipping routes.

d. a dynamic data file containing information on merchant and military surface and subsurface traffic volumes, and strategic resource movements.

e. a remote sensing suitability index file which will be loaded as data becomes available. This index will be designed to identify areas suitable for the collection of required hydrographic data utilizing remote sensing technology.

Cell Structure

HYSUR will broaden the concept of the Chubb Study's mission area from 6 - 300 fathom depths to 0 - 1000 fathom depths in order to accommodate all civil and military requirements. Mission areas will be subdivided into smaller areas called cells based on depth, geomorphological characteristics, and military and civilian usage. Cells will be categorized as Coastal Cells (with depths from 6 - 1000 fathoms along one body of land), Strait Cells (with depths from 6 - 1000 fathoms, located between two physically close land areas and through which maritime traffic will occur on each land side), Harbor Cells (with depths from 0 - 1000 fathoms in which vessels may dock or anchor for cargo loading or transfer), and Amphibious Assault/Combat Cells (coastal areas with depths from 0 - 1000 fathoms designated for military operations).

After the mission areas of the world are segmented into appropriate cells, the size of each cell will be measured using a planimeter on small scale world charts overlayed with the appropriate cell structure. Cell sizes will be loaded into the static data file, and will be available as an additional factor to determine the application of survey resources.

Identification

HYSUR will access information on hydrographic source coverage for each desired cell area through interfaces with several existing DMA data bases (such as, the Area Requirements And Product Status System (ARAPS) and the Bathymetric Information System (BIS)) and HYSUR data files. ARAPS will provide information on area requirements including priority, production status, chart evaluation, and intended use. BIS will provide information on bathymetric surveys. HYSUR will automatically identify source data deficiencies and highlight needed survey areas.

Value

A numerical ranking value will be derived for each cell based upon the characteristics of actual and potential civilian and military usage for each cell. A cell with a high rank value has a higher survey priority than a cell with a lower rank value. The following factors are involved in the computation of the cell weight:

- Wt: Cell Weight
- SR: Strategic Resource Weight (range 0 to 100)
- MT: Merchant Shipping Weight
- S : Submarine Operations Weight (range 0 to 99)
- B : Mine Warfare Weight (range 0 to 99)
- J : Mapping, Charting, and Geodesy (MC&G) Priority (range 1 to 5)

This relationship is expressed in the following equation (Eq. 1):

$$Wt = SR + \frac{MT + S + B}{J} \qquad (1)$$

Strategic Resource Weight. The strategic resource weight, SR, represents the percentage of U.S. imports of all strategic materials on designated shipping routes. For example, if 92% of all commodities transported on a specified shipping route is delivered to the U.S.*, then SR is assigned a value of 92.

Merchant Shipping Weight. The merchant shipping weight, MT, is expressed as follows (Eq. 2):

$$MT = WT \times (1 - dm/wm) \qquad (2)$$

where,

- WT: Importance factor of port or route (range 0-100)
- dm: Distance in nautical miles from the 20-1000 fathom depths
- wm: Distance in nautical miles from the 6-1000 fathom depths
- (1-dm/wm): Degree of safe passage for merchant vessels
  - $(1 - dm/wm) \rightarrow 0$: Relatively safe passage
  - $(1 - dm/wm) \rightarrow 1$: Limited safe passage

Equation 2 portrays the risk factor for important merchant shipping through certain areas. The greater the risk, the higher the merchant weight.

The importance factor of a port or route, WT, is derived as follows (Eq. 3):

$$WT = WC \times WV \qquad (3)$$

where,

- WC: Strategic importance of cargo, derived from the commodity code (range 0-10)
- WV: Factor derived from the volume of cargo

Equation 3 quantifies the importance of the port or route by relating the volume and type of cargo, thereby preserving the significance of small shipments of strategic materials.

Strategic Importance. The strategic importance, WC, is a number assigned to each strategic commodity according to its importance to the U.S. It ranges from 0 (no strategic importance) to 10 (great strategic importance).

Cargo Volume. The optimum basis for measuring merchant trade is generally considered to be by volume or tonnage. However, the only readily available comprehensive base of statistics for world trade was found to be expressed as cargo value in U.S. dollars. The cargo value has a wide range, i.e., from $1 million to $1 trillion. In order to effectively weight this large range of data, cargo volume, WV, is expressed logarithmically (e.g., a cargo value of $100 million reflects a cargo volume WV = 1, for $1 billion WV = 11; and for $1 trillion, WV = 31).

---

*Extracted from U.S Oceanborne Trade Statistics and the Yearbook of International Trade Statistics, and will be contained in the dynamic data file.

The following three examples demonstrate the computation of merchant shipping weight:

Example 1 - A cell with $1 billion (WV = 11) worth of platinum cargo (WC = 4) will have a route importance factor, WT, of 44.

WT = WC x WV = 4 x 11 = 44

Example 2 - If the cargo in Example 1 is routed through a cell where the distance from the 20 fathom depth to the 1000 fathom depth is 150 nautical miles, and the distance between the 6 fathom depth to the 1000 fathom depth is 450 nautical miles, that cell will have a merchant weight of 29.

MT = WT x (1 - dm/wm)
From example 1, WT = 44
WT = WC x WV = 11 x 4 =44

Therefore,

MT = 44 x (1 - 150/450) = 29

Example 3 - If the cargo in Example 1 must pass through a cell with 150 nautical miles between the 20 fathom depth and the 1000 fathom depth, and 300 nautical miles between the 6 fathom and the 1000 fathom depth, the merchant weight is 22.

MT = 44 x (1 - 150/300) = 22

Since Example 2 expresses a higher navigational risk, 2/3, than Example 3, 1/2, the area in Example 2 has a higher merchant shipping weight.

Submarine Operations Weight. The submarine operations weight, S, is expressed as follows (Eq. 4):

$$S = WS \times (1 - ds/ws) \qquad (4)$$

where,

WS: Submarine weighting factor (0 or 99)
ds: Distance in nautical miles from the 300 - 1000 fathom depths
ws: Distance in nautical miles from the 20 - 1000 fathom depths
(1 - ds/ws): Degree of safe passage for submarines
(1 - ds/ws) → 0: Relatively safe passage
(1 - ds/ws) → 1: Limited safe passage

If a cell area is designated as a possible submarine interest area, the submarine weighting factor is 99; if not, 0 is applied to that cell. The degree of safe passage for submarines is analogous to that for merchant shipping.

Mine Warfare Weight. The mine warfare weight, B, is derived as follows (Eq. 5):

$$B = WB \times (1 - db/wb) \quad (5)$$

where,

- WB: Mine warfare factor (0 or 99)
- db: Distance along the narrowest passage with a depth of 1000 fathoms
- wb: Total length of the passage
- (1 - db/wb): Degree that traffic may be blocked
  - $(1-db/wb) \rightarrow 0$: minimum blockage attainable
  - $(1-db/wb) \rightarrow 1$: maximum blockage attainable

If a cell area is designated as a potential mine warfare area, the mine warfare factor of 99 is assigned; if not, a factor of 0 is applied to that cell.

Example 4 - A cell is designated as a potential mine warfare area. The distance along the narrowest passage with a depth of 1000 fathoms is 30 nautical miles and the total length of the passage is 300 nautical miles. The mine warfare weight for this cell is 89.

$$B = WB \times (1 - db/wb) = 99 \times (1 - 30/300) = 89$$

MC&G Priority. The MC&G priority, J, is specified in the Joint Strategic Planning Document (JSPD). The priority is expressed as a number ranging from 1 (the highest priority) to 5 (the lowest priority) and is assigned by geographic area. Since the merchant, submarine operation, and mine warfare weights will be divided by the MC&G priority value, MC&G priority plays a primary role in assigning a weight to each cell.

Example 5 - A cell containing an established shipping route on which 92% of commodity trade is imported by the U.S. has the same merchant weight as shown in Example 2. If a submarine operation weight of 99, a mine warfare weight of 0, and a hypothetical MC&G priority of 2 is assigned to that cell, then the cell weight will equal 152.5.

$$Wt = SR + \frac{MT + S + B}{J}$$

$$Wt = 92 + \frac{22 + 99 + 0}{2} = 152.5$$

Example 6 - If the cell in Example 5 has an MC&G priority of 1, Wt is increased to 213.

$$Wt = 92 + \frac{22 + 99 + 0}{1} = 213$$

Weighting factors can be used in any combination. For instance, one or more weighting factors can be deleted or any weighting factor can be increased during calculation to suit a user's specific purpose.

Planning

After a cell is assigned a numerical value based on the weighting factor, a survey precedence is established. The product requirement (type of chart, scale, geographical position, intended use, etc.), data deficiencies, urgency of the requirement, political accessibility, bilateral agreements, as well as the weighting factors for a cell are utilized in the survey area ranking procedure. To minimize transit times, the system weights all cells and then groups the cells by geographic proximity. A Survey Requirement Planning Report in the form of textual information in soft or hardcopy form will then be produced. The system will also produce a magnetic tape for hardcopy plot generation. This graphic plot will portray the mission area, cell area, priority, shoreline, DMA chart coverage, other source coverage, volume of merchant and military traffic, planned and in-process surveys, strategic resource movements, and ship routes.

CONCLUSION

The potential benefits from this planned system, when it is fully operational in 1985, are that DMA and NAVOCEANO planners can make more effective and timely decisions by virtually eliminating the intensive manual and semi-automated activities to determine priorities for required hydrographic surveys. The cost of operating a survey ship, approximately $30,000 per day, accentuates the potential savings of an effective management planning system like HYSUR. Even more important, however, is the resulting capability to apply limited resources to satisfy the most critical survey requirements. Thus, more effective use of extremely limited resources can be achieved. Rapid deployment programs and crisis areas can also be more easily supported by allowing ready retrieval of source information for decision making.

Continuing emphasis must be given to other ways of acquiring needed hydrographic data. With rapidly changing technology, hydrographic data acquisition will not depend solely on existing conventional surveys. Utilization of space sensors, such as multispectral scanning systems, synthetic aperture radars, thermal band scanners, or airborne sensors (e.g., the Hydrographic Airborne Laser Sounder) will be employed. HYSUR will be designed to accommodate these systems.

COMPUTER-ASSISTED CARTOGRAPHY FOR CENSUS COLLECTION: CANADIAN ACHIEVEMENTS AND CHALLENGES

Joel Z. Yan
Geocartographics Subdivision
Data Processing Division
D. Ross Bradley
Geography Division
Statistics Canada
Ottawa, Canada, K1A 0T6

ABSTRACT

To ensure an orderly coverage of the entire land mass of a country, Census enumerators are given maps which provide precise descriptions of their spatial areas of responsibility. The Census taking agencies in Canada and the United States have launched development programs which for subsequent Censuses will lead to large numbers of these collection maps being produced by computer-assisted means. This paper focuses on the Canadian experience which is based on results from the 1981 Census of Canada (34 maps produced over three months using a prototype system) and from the volume testing of the production system (200 maps produced in one week) being implemented for the 1986 Census. This production system is currently being enhanced with the aim of producing approximately 1200 collection maps for the 1986 Census. In addition to presenting the achievements to date, this paper discusses the challenges and opportunities which remain.

1. INTRODUCTION

To carry out a household survey based on area sampling methodology, good, easy to read, large scale maps are necessary to ensure that enumerators or interviewers know exactly the geographic area that must be covered in the collection of data. This statement is true whether it be a sample survey such as the Canadian Labour Force Survey (56,000 households) conducted on a monthly basis or whether it be the Census of Canada (9,300,000 households) with its objective of counting every person usually resident in Canada at his or her usual place of residence on Census Day. To provide these maps is a lengthy, costly Statistics Canada operation carried out by a team of clerical staff and draftspersons, which required approximately 80 person years and $1,500,000 for the 1981 Census. For several years the Census has been investigating the use of automated mapping techniques to reduce the cost of this operation and to improve the consistency of map bases used for the collection, processing, and retrieval of census data. Meanwhile, the U.S. Bureau of the Census is also planning to do its collection mapping by automated means for the 1990 Census. This paper highlights the development of the Canadian system from 1979 to the present.

2. DEVELOPMENT AND EVALUATION OF A PROTOTYPE SYSTEM 1979-1982

2.1 Objective

In the fall of 1979, a study was authorized by senior management at Statistics Canada to explore the feasibility of producing, on a highly selective basis, census collection maps by semi-automated means. They were to be tested for operational adequacy during the collection of the 1981 Census. The study was initiated in the interest of saving human and financial resources through the use of technology and the increased integration of activities. The results of this study have been reported in detail elsewhere (Bradley 1981; Yan, 1982), and only a brief summary is provided here.

2.2 Scope

Three separate areas were chosen. They were (i) a small portion of the City of Kingston, Ontario, (ii) a large section of the city of Gatineau in Quebec, and, (iii) the entire Federal Electoral District of Windsor West in Ontario. These areas were chosen because of their proximity to Ottawa, the familiarity of the areas by the staff who were to work on this task, and the good quality of the Area Master File from which this information would be retrieved.

2.3 The Area Master File

Area Master Files contain a logical representation of all city streets and other geographic features such as railroad tracks, rivers, and municipal boundaries in machine readable form. The Area Master File (A.M.F.) is a product of the Canadian Geocoding system which was introduced more than 10 years ago to enable retrieval of census data for user specified areas. The AMF is maintained by the Geocoding Unit of Geography Division in Social Statistics Field. It corresponds in function, to the GBF/DIME file created in the U.S. during the same period. There are differences in structure since the DIME file is based on the block, and the AMF is based on the blockface.

Large urban areas (population 50,000 and over) are divided into block faces. A block face consists of one side of a street between two successive intersections. These block face spatial units are small enough that when aggregated they become a good approximation for a user identified query area. Each block face is assigned central x-y UTM coordinates, to which files of households, or persons can be coded i.e., geocoded. A user needing information from a geocoded file, outlines his/her area of interest on a map. This area is digitized and becomes a special "query area". All block face centroids falling within this area are aggregated and statistical data from the Census are tabulated for those block faces. This process is described in the booklet "Facts by Small Areas" (Statistics Canada, 1972). New applications of the AMF have been described by Boisvenue & Parenteau (1982). Area Master Files now exist for virtually all urbanized areas of the 36 tracted centres in Canada of 50,000 population and over. This constitutes coverage of over 60% of the Canadian population.

2.4 Computer-Assisted Mapping System

These days, most producers of computer-assisted maps enter their digital information in a form appropriate for cartographic representation. For example, text locations, orientations, and sizes are defined explicitly at data entry time. Maps can be produced, but there is usually some difficulty in deriving a geographic base file where the texts are related to the appropriate point, line or area feature. Similarly it may be difficult to derive the network or graph formed by the various linear features between intersections. Most federal and state or provincial mapping, agencies would fall in this category. Some cities such as Calgary are building combined cartographic and geographic base files, but this is the exception. In the case of the Canadian and U.S. census agencies, geographic base files existed with their network structure, but not urban cartographic base files with explicit text locations, and highly developed feature symbolization defined.

In order to generate double line street census tract collection maps according to specifications a number of new capacities had to be developed which were not explicit in the AMF. These included:

(i) subdividing a single AMF (which covered one or more municipalities) into subfiles by census tract (CT) or other query polygon;
(ii) the formation and shrinking of block boundaries from AMF features to leave a double line street pattern;
(iii) a reasonably intelligent assignment and placement of block numbers within census tracts;
(iv) sensible placement of feature names and corner civic addresses; (see Figure 2)
(v) generation and plotting of standard cartographic features such as north arrow, title block, scale bar.

System development activity was undertaken late in 1979. Given the short time frame available, it was not possible to develop a complete production system for the 1981 census. Instead, a prototype system was built by enhancing existing software. Blocks were formed and shrunk on an Amdahl V8 computer to create double line street patterns. The AMF's were then subdivided or "clipped" by polygon and street names and addresses were positioned on the base map using a minicomputer. Thereafter, street names were shifted interactively and additional text not resident on the input file was added interactively using the AUTOMAP cartographic system. Three or four cycles of text editing and plotting were required to produce a final quality map. On the average, 2.2 hours of interactive editing were required per CT map.

2.5 Results

An evaluation and cost benefit study was conducted for the 32 CT maps and 200 enumerator maps prepared by computer-assisted means (see Yan, 1982). The maps were found to be more than operationally adequate in quality. Even with the high interactive edit times, the costs of producing the maps from the AMF for the prototype system were similar to the costs of manual drafting for preparation of the base maps. Furthermore, the average time per map was reduced from 17 to 9 person-hours. However, due to lengthy plotting times the cost for updating the base map and producing the final CT diagram remained less with the traditional method. The actual dollar costs and number of person hours (PH) measured with this prototype system are given below.

Table 1. The Average Cost to Produce a CT Collection Map

	TRADITIONAL METHOD		PROTOTYPE COMPUTER-ASSISTED METHOD	
Preparing the CT Base	$210	17.0 PH	$181	9.0 PH
Updating the Base	$ 65		-	
Producing the diagram	$ 18	0.7 PH	$100	4.5 PH

A further cost saving of 30 to 40 per cent was predicted with development of a production system for computer-assisted mapping. Two other benefits were expected from the computer-assisted approach. Firstly, a savings in the geocoding census capture operation was expected given that the CT map used for collection would be consistent with the A.M.F., which is used for processing and retrieval of the census data. Secondly, computer-assisted mapping could eliminate the twofold updating operation of street patterns: on the CT maps and on Area Master Files.

Given these anticipated benefits, the following decisions were made in April of 1982:

(1) Develop and implement a production system for computer-assisted collection mapping to reduce the costs and manpower requirement of the current prototype system.
(2) Produce a large number of base maps for the 1986 Census by the computer-assisted method, including all maps requiring re-drafting and all maps for new tracted centres.
(3) Maintain the traditional method of producing and updating collection maps as backup and implement work-sharing with the computer-assisted system.
(4) Move towards utilizing the AMF as the unique base for urban areas and extend it to cover the entire urbanized core of cities of population 50,000 and over.
(5) Extend the cartographic content of the AMF base to include additional features for collection mapping.
(6) Investigate potential links with other areas in Statistics Canada, and other government departments and agencies for automated mapping production methods.
(7) Review these decisions in early 1983 based on the results of application experiences.

3. DEVELOPMENT OF A PRODUCTION SYSTEM 1982-1983

3.1 Objective

The development of a production system was undertaken to reduce the cost and increase the throughput and reliability of the prototype system described above. The production system will be utilized to produce 1200 CT maps encompassing on the order of 8000 enumeration area maps for the 1986 Census. These maps must be provided for verification in the field by mid 1985.

3.2 Improvement to the Prototype System

In May 1982 development of the production system began in earnest. Figure 1 shows an overview of the system developed.* Major enhancements which were included are described below.

3.2.1 Block Formation The system was completely written to use "incore" processing as compared to disk I/O. The result was a 100-

* The authors wish to acknowledge and express their appreciation to the System development team including John Crawford, Joseph Gomboc, Ron Cunningham, David Tupper, Robert Porteous, Cathy Gourley and Gaetan St.-Pierre. Sid Witiuk, Andre Boisvenue, Gilles Gavard and students from the CEGEP of Hull are also acknowledged for initial work on the block formation system.

fold increase in speed and corresponding reduction in price of block formation. The final system retains the initial approach for double line street pattern generation of forming a block and then shrinking it. Van Est (1982) has instead used a segment intersection method for generating the double line street pattern. This allows for variable width streets but not for blocks as entities themselves.

3.2.2 **Polygon Clipping** The clipping program was rewritten for the main frame computer (Amdahl V8). It became the first step in the process, utilized to subdivide a large A.M.F. into more manageable units before block formation.

3.2.3 **Cartographic Prompter** An online prompter program was developed to review the output from clipping, and to query for cartographic control information including map scale, and map title information.

3.2.4 **Model Loading and Plotting** This system was retained on the HP-1000 mini-computer in order to facilitate on-line viewing and editing of the maps. However, a much streamlined batch procedure was developed which greatly reduced the requirement for operator intervention. The result was an 80% reduction in the cost of processing and connect time from approximately $20 to $4 per map. At the same time, an improved text placement system was developed. The objective was to produce acceptable quality CT maps without on-line cartographic editing which had consumed 2.2 hours per map with the prototype system. Consultation with cartographers and draftspersons played a useful part in this enhancement. Some of the major elements included:

(a) plotting the name centered along features every eight inches based on the final plot scale.
(b) breaking the name into composite words and attempting to fit each word into a street segment rather than the complete name.
(c) ensuring that names of dead end streets do not cross over the junction streets.
(d) keeping track of the number of names which do not fit and coding their names for plotting at small size and perhaps for online displacement.
(e) providing the option of plotting addresses at block face centroids or street intersections.
(f) automatic assignment of block number, block number size and location within the block.

3.2.5 **Map Touch-up** With the initial requirement that CT maps fit on a single plotter sheet (70 x 100 centimeters), the textual and symbolic annotation could be very dense. It was recognized that some human intervention or "map touchup" system would be required. Two types of touch-up were expected: displacement of text to prevent overlap, and addition of cartographic information not currently contained on the digital input files. These changes could usually be made manually on the final plotted maps, but given the need to plot such maps at regular intervals, (at least twice per census), an automated touch-up which could be made once digitally, and utilized subsequently seemed preferable. But, under what circumstances were touch-ups required? To examine this issue further, it was decided that a volume test should be conducted.

4. THE VOLUME PILOT STUDY

4.1 Objectives

The major objectives of the volume pilot study were to measure the progress made in the system development activity and to determine further developments and improvements required to ensure an effective and efficient production system.

4.2. Scope

The test was conducted over a two week period in January with the goal to produce as many maps as possible in that time. The test involved utilizing the system as it existed with no manual intervention or touch-up. A detailed map checking procedure was developed involving careful review of the maps by three groups: draftspersons familiar with the specifications required, the system team, and AMF maintenance personnel.

4.3 Volume Test Results

Detailed results have been reported elsewhere by Yan (1983) and Bradley (1983). The developed system performed well.

4.3.1 Throughput 207 CT edit maps were produced in 6 work shifts. Throughput was approximately:

> 45 CT maps per shift through the prompting and block formation phase and 32 CT maps per shift through the model loading, text placement, and plotting phase.

Throughput for production of the final edited mylar maps is expected to be at least 10 maps per shift.

4.3.2 Costs Based on the 207 maps processed, the average production cost for an edit plot on paper was approximately $13. With provision for checking, correction, reprocessing and plotting on mylar the total expected production cost per base map is expected to be approximately $60, compared to $210 for redrafting a map by the traditional method. Significant reductions in person hours were also experienced.

4.3.3 Quality of Outputs The quality of the maps improved significantly from earlier tests, but still required further amelioration. In general, automatic text placement in the urban core CT maps at scales of 1:2400, and 1:4800 was quite good (see Figure 2), but the CT maps at scales of 1:7200 and up were too crowded for the text to be clear.

4.3.4 Required Improvements From a detailed check of the volume test maps, a list of problems was drawn up. There were nineteen problem areas identified related to the input boundary and street network files, and fifteen problem areaa related to the double line-mapping system itself. Analysis and resolution of the problems is underway.

Most of the problems with the inputs arise from the fact that the

AMF was designed and created primarily as a geographic base for georeferencing rather than as a multi-purpose cartographic base. Some issues such as the addition of parks, and railway spur lines are being handled easily within the current framework. Other cartographic shortcomings identified are being handled by increasing the repertoire of feature types to uniquely identify highway ramps, cliffs, fences, divided roads etc. It appears that in some cases a location for plotting a feature name may have to be stored as well as the coordinates which define the feature itself. Other issues such as the addition of the French accents on names require more detailed analysis.

Most of the system-required enhancements were related to the automatic placement of text by the system. Many streets were too short for the complete name to fit at the plot scale requested. In other cases, the names of nearby features overlapped. In Quebec City, for for example, 42% of the names did not fit when plotted at 1:4800 scale, whereas only 16% did not fit when plotted at 1:2400. In the future, the plot scale will be reviewed based on the number of names which "appear to fit". Some of the decisions taken which have already improved the quality of text placement since the volume test include:

(a) leaving dead end streets open so the street name will not overlap the line at end of the street;
(b) shifting text slightly above most singleline features (e.g., railways, creeks) but slightly to the right of boundary type features (e.g., city limit names) to minimize the chance of text overlap;
(c) checking for adjacent segments of similar slope for plotting longer names;
(d) leaving names off railway yards, highway names and other features where the name is not required.

Resolution of these problems is well underway and significant progress is expected before the final maps are delivered for the 1986 Census.

4.3.5 <u>Lessons Learned</u> Some of the lessons learned which could be applied to others developing computer-assisted mapping applications are enumerated below.

Lesson 1: Utilize the automated system for what it does best. Do not insist that the system produce the final map completely, but utilize the "human system" for the typical touchups that aren't yet in digital form.

Lesson 2: Set up a project team which includes experienced personnel from both drafting/cartography and systems development and ensure a high level of communication, with the cartographers providing the map specifications.

Lesson 3: Keep the system friendly, responsive, and easy to use, so that it will more likely be viewed as a tool and not as an obstacle by drafting personnel.

5. ACHIEVEMENTS AND PREDICTIONS

The Canadian achievements to date in the area of computer-assisted collection mapping have been notable:

(i) a high throughput of map production has been demonstrated in the volume test; (32 edit maps/shift)
(ii) significant production cost savings have been achieved in generating a map from an existing base file as compared to redrafting; ($60 compared to $210)
(iii) sample map products for other applications such as Canada Post and Labour Force Survey have been produced;
(iv) work has begun on investigating digital data exchange with agencies such as Energy Mines and Resources Canada, (DIGITOP Program), provincial governments (Quebec), and local governments (cities of Calgary, St. Catharines, and Winnipeg).

American achievements in this area have included production of sample enumerator maps for the 1980 Census and ambitious plans for the 1990 Census. More details are reported by Broome (1983) elsewhere at this conference.

5.1 Throughput and Cost Savings

If the current throughput of 32 maps/shift were retained all 1200 base maps could be produced within 40 working shifts (8 weeks). Given the additional steps of editng, reprocessing and plotting on mylar to achieve final quality, this is expected to require approximately 3 hours per map of computer operator/draftsperson. This compares with approximately 17 hours per map by the traditional method.

Assuming a final production cost of $60 per map, the computer-assisted mapping system should provide a savings of approximately $150 per map, compared to drafting maps by the traditional method. The savings for 1200 maps of approximately $180,000 will be applied to the development cost for a production system. On the assumption that a cost effective touch-up system can be developed, and that the input files can be extended in cartographic content, the future of computer-assisted collection mapping for the Census of Canada will be bright. If production of the 1200 maps goes well for the 1986 Census, the number of census collection maps produced by the computer-assisted method is likely to increase substantially for the 1991 Census.

5.2. Other Mapping Applications

Since the system has been developed taking into account a generalized set of requirements, there are applications to other areas besides Census. The system has been designed to produce a set of maps to user specifications from any AMF, and any set of polygon boundaries. A number of different applications have recently been demonstrated using a prototype data linkage system including:

- CT and Forward Sortation Area (F.S.A.) double line maps with postal code shown at block face centroid for Canada Post Corporation; (see Figure 3)
- CT maps showing number of voters per block face to meet the election operations needs of the Chief Electoral Officer of Canada;
- CT maps showing data variables such as household and population count per block face for field collection purposes.

In the future, one could imagine that a user could request an urban computer-assisted map tailored to his/her specifications. He or she could specify the map scale, title, and layout and select from a menu that included:

- the polygons or study areas of interest
- the urban area or AMF to be processed
- the census data variable to be displayed by block or block face (as long as data confidentiality was respected)
- single or double line street patterns
- block face address ranges
- other cartographic features.

6. REMAINING CHALLENGES

The future for computer assisted mapping appears to be bright, but several challenges remain. Technology would appear to be in place to allow for the production of large volumes of maps for the specific purpose of collecting information in an orderly manner. Costs for putting a system in place are not low, but the benefits which can accrue will probably render these costs most affordable. The important factor to be considered is that many institutions should be in a position to take advantage of this technology. Their needs, while different, have many striking similarities. The Labour Force Survey and the Census of Canada collect information from households for some of the most sought after statistical series in Canada. Enumerations for election purposes perform a similar function to those of the Census. Canada Post, in carrying out its mandate to deliver the mail, requires a map for control and administrative purposes. The same can be said for municipalities as they provide the many essential services, especially in large urban areas. There has been significant progress in computer-assisted collection mapping but several major challenges remain.

<u>To the computer-assisted mapping team</u>:

(1) complete the development which has begun for the 1986 Census of Canada and the U.S. 1990 Census to provide efficient production of quality maps.

<u>To the managers of mapping agencies</u>:

(1) organize and coordinate the needs among major users to enhance compatibility and reduce overall development and production costs.
(2) work towards the multi-purpose cadastre.

<u>To system developers</u>:

(1) develop data base structures and organizations which can support the multi-purpose cartographic requirements and integrate the various existing data sources;
(2) resolve some of the more difficult technical problems such as automatic text placement and data generalization;
(3) develop effective methods of updating the geographic and cartographic base through interactive technology and from linkage to external digital files.

<u>To cartographers</u>:

(1) take a more active part in the specification and development of the automated cartography systems of the future.

The task of coordinating these needs is formidable, but one which should be undertaken. There must be cooperation among the many users and developers so that the expensive systems can be afforded by some method of cost sharing. Unnecessary duplication of effort can no longer be afforded.

REFERENCES

1. GRDSR: Facts by Small Areas, Statistics Canada, June 1972.
2. Boisvenue, A. and Parenteau, R., "The Geocoding System in Canada and its Area Master File", in Papers from the Annual URISA Conference, August 1982, pp 226-232.
3. Bradley, D.R., "Computer Assisted Collection Maps for the 1981 Census of Canada", proceedings of the National Advisory Committee on Control Surveys and Mapping Seminar on the "State and Utility of Computer-Assisted Map Production", November 1981.
4. Bradley, D.R., "Report to Senior Management on Results of the CACM Volume test", Statistics Canada, May, 1983.
5. Broome, F.R., et.al.,"Longterm Plans for the Geographic and Cartographic Support to the U.S. Bureau of the Census", Proceedings of Auto CARTO VI.
6. Van Est, J., DeVroege, F., "The meaning and construction of a geographic base file for a spatially oriented information system", European Symposium of Urban Data Management, Valencia, Spain, 1982.
7. Yan, J.Z., "Evaluation Report of Computer-Assisted Collection Maps for Census", 44 pp., Statistics Canada, March 1982.
8. Yan, J.Z., "Computer-Assisted Collection Mapping for Census: Application Experiences", Proceedings of Graphics Interface '83, Edmonton, May 1983.

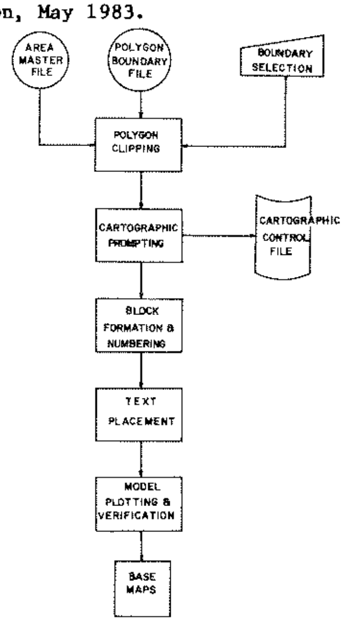


Fig. 1 Flowchart of the Computer-Assisted Collection Mapping System at the Census of Canada.

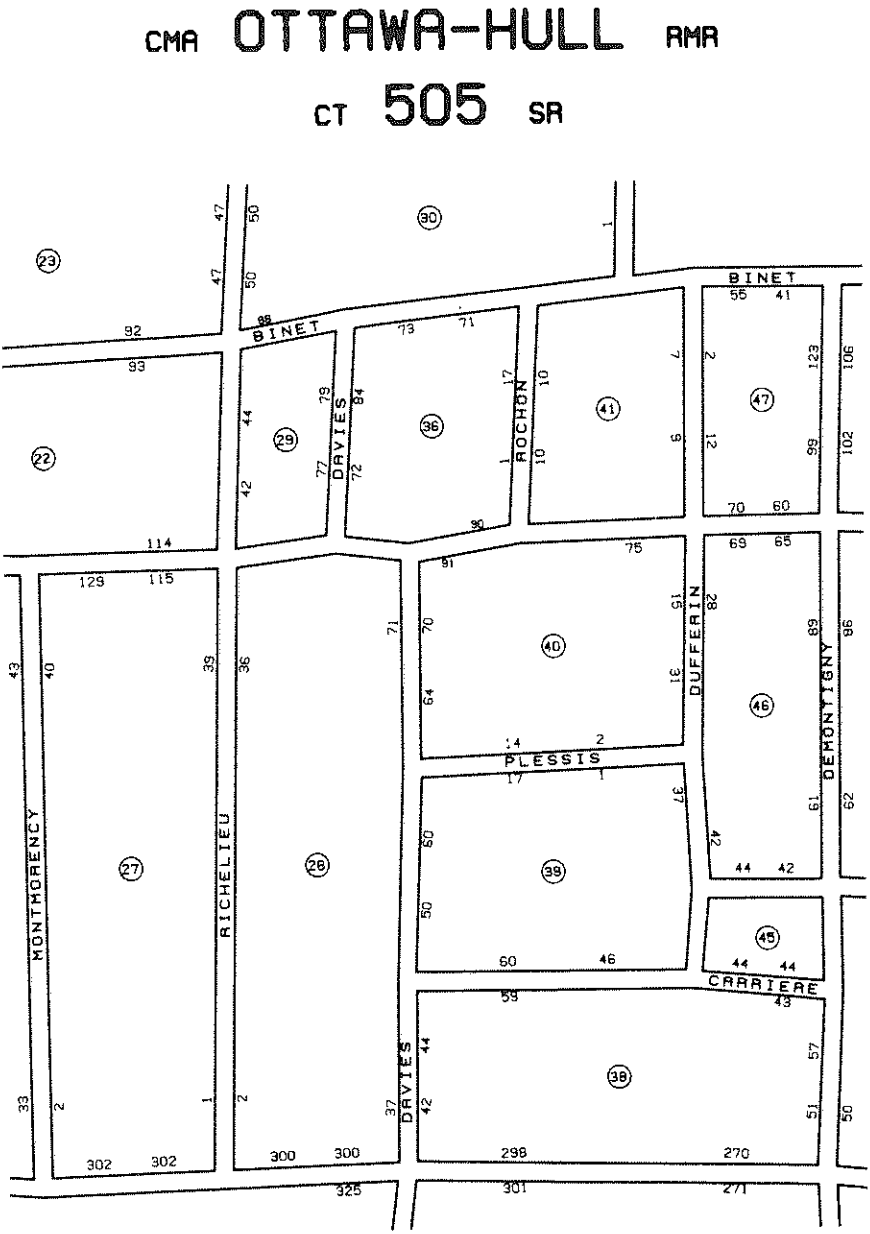


Fig. 2 Part of a census tract map showing address ranges and block numbers produced by automated means for census collection. There has been no manual intervention.

CMA **OTTAWA-HULL** RMR

CT **505** SR

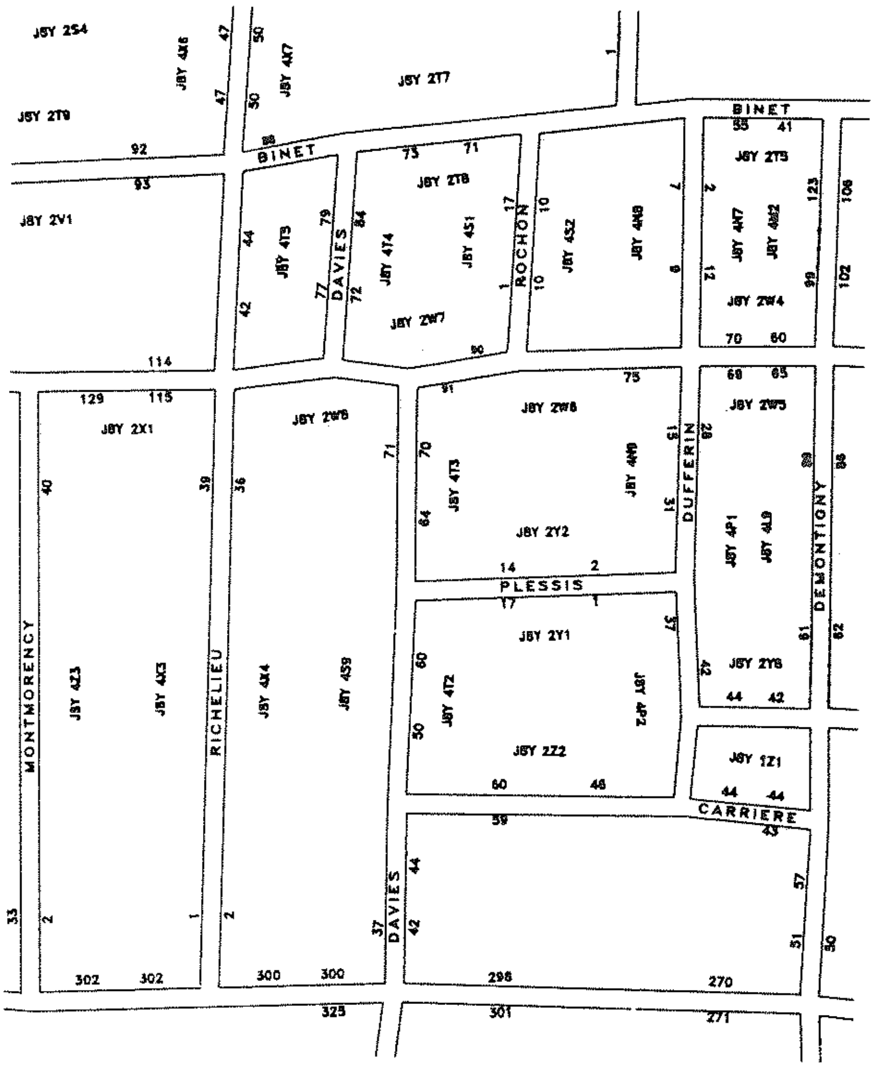


Fig 3 Part of a census tract map showing address ranges and postal code produced by automated means.

L'UTILISATION DES SYSTEMES GRAPHIQUES INTERACTIFS EN CARTOGRAPHIE

Dominique Bruger
Service de la cartographie
Ministère de l'Energie et des Ressources
1995, bl. Charest Ouest
Ste-Foy, Québec
G1N 4H9

RESUME

Entre 1978 et 1981, le Service de la cartographie du ministère de l'Energie et des Ressources du Québec a effectué des études de faisabilité avec un système graphique interactif de capacité réduite. Depuis septembre 1981, le ministère a commencé la production mumérique des nouvelles cartes photogrammétriques et la création d'un fichier des limites administratives de la province. La production numérique des cartes photogrammétriques est divisée en trois étapes: le captage ou saisie de l'information, l'édition qui correspond à la symbolisation cartographique et le traçage. Le ministère a confié la production à l'entreprise privée et travaille en étroite collaboration avec elle pour établir des spécifications et des développements informatiques rentables. Le ministère travaille également à la préparation d'un fichier des limites municipales du Québec associé à un fichier de données alphanumériques. Le fichier graphique a été créé en numérisant les limites compilées manuellement sur des cartes 1/20 000 ou 1/50 000. Dans le fichier graphique, chaque municipalité est représentée par un polygone fermé. Le fichier alphanumérique contient des informations fournies par le Bureau de la Statistique du Québec. Certains éléments des cartes numériques 1/20 000 et le fichier des limites municipales constitueront des données initiales pour la base géographique québécoise.

ABSTRACT

Between 1978 and 1981, the Service de la cartographie of the Ministère de l'Energie et des Ressources du Québec has driven a serie of feasibility studies with a reduced capacity interactive graphic system. Since September 1981, the ministère went into production. Until now, the digital method is used for producing new photogrammetric maps and for creating a graphic file containing Québec's administrative limits.

The production of digital photogrammetric maps is divided into three steps:

- information collection. The data is taken directly from air photographs and can be viewed instantly on a graphic terminal screen. The data file contains gross information, with exact positioning but with no symbolization. The graphic elements produced are not coded, due to the fact that the production process is not oriented towards a data base, although those gra-

phic elements are organized in a way that permits further sorting with the aid of programmed techniques.
- editing phase, which consists of transformation made on the collection filè to give it a cartographic representation symbology.
- plotting phase, made with a high precision, automatic drafting table.

The ministère conserves the same management politics with digital mapping as those applied to conventional mapping, this means that the work is given to private enterprises. In this particular case, a 5 year contract was granted to a consortium composed of a private company and a Crown Corporation. At the beginning, the ministère and the Consortium were put into close contribution to establish the new specifications and to develop the necessary add-on software. In a second phase, the ministère worked on methods to increase the productivity.

The ministère is also working in the conception and creation of a graphic file containing the municipal boundaries for the whole Québec province. The objective is to obtain cartographic representations resulting from a certain statistical sorting. The file is divided in two subfiles: the graphic file and the attributes file. The graphic filë is created using manually compiled limits on 1/20 000 and 1/50 000 base maps, which are then digitized. The provincial graphic filè uses geographic coordinates (latitude and longitude) and represents each municipality as a closed polygon. The attributes file contains information concerning those municipalities, given by the Bureau de la Statistique du Québec. The municipality centroïd in the graphic file is associated by software to the attributes file: the link between the two filès is thus insured.

After two years in production, the photogrammetric digital mapping has grown to it maximum degree of productivity. With the first 1/20 000 digital maps, the ministère is well equipped to put his foot into the field of geographic data bases, of which our present research activities consist. The 1/1000 digital maps allow the creation of integrated digital systems for municipal management.

INTRODUCTION

Le ministère de l'Energie et des Ressources de la province de Québec est responsable de la production de cartes photogrammétriques à grande échelle et à moyenne échelle (1:1 000 à 1:20 000), de certaines cartes thématiques ainsi que de cartes générales de la province à très petite échelle (1:1 000 000 à 1:9 000 000).

En 1978, le Service de la cartographie a fait l'acquisition d'un système graphique interactif de dimensions réduites. Ce système était destiné à analyser le potentiel de la technologie numérique appliquée à la carte photogrammétrique. Après deux ans et demie de formation, de tests et de développement sur ce système, le ministère a décidé d'amorcer l'étape de la production.

La présente communication décrit les caractéristiques des travaux de cartographie numérique en cours de réalisation depuis septembre 1981 ainsi que les objectifs poursuivis par le ministère.

ORIENTATION DE LA PRODUCTION NUMERIQUE

La technologie numérique se présente sous différents aspects, et conduit à des méthodes sensiblement diverses, dépendant du choix des intrants (informations à numériser) ou des extrants (produits désirés).

En cartographie, les principales sources d'information sont;
- les photographies aériennes
- les cartes topographiques existantes
- les informations thématiques compilées graphiquement
- les levés topographiques

Les sous-produits des données numériques peuvent être également très variés:
- produits graphiques diversifiés répondant aux besoins de la clientèle
- production de bases de données
- sorties statistiques alphanumériques

Actuellement, la production des cartes topographiques constitue le mandat le plus important du Service de la cartographie, même si le domaine des cartes thématiques élargit ses champs d'activités. Le ministère a donc décidé d'appliquer la technologie numérique à la production des cartes de base. Par ailleurs, il était plus logique et plus rapide de travailler sur des cartes nouvelles, plutôt que sur la numérisation de cartes existantes. Le ministère a donc opté pour des projets de cartes photogrammétriques nouvelles, numérisées directement à partir des photographies aériennes.

Environ un an après l'amorce de la production numérique des cartes photogrammétriques, nous avons abordé la question de la cartographie thématique c'est-à-dire la numérisation d'informations graphiques compilées manuellement sur des cartes existantes. Ce premier projet de cartographie thématique s'est concrétisé par la création d'un fichier des limites municipales du Québec.

Alors que la sélection des intrants n'est en fait qu'un ordre de priorité, il en va tout autrement pour le choix des produits. En effet, cette décision a une grande influence sur la rentabilité des projets et sur l'influence et la respectabilité de la nouvelle technologie à moyen terme.

La production numérique a été résolument orientée vers les sous-produits graphiques pour deux raisons majeures.

D'une part, pour l'immense majorité de la clientèle, le produit graphique restera toujours le document de travail privilégié. La technologie numérique possède une grande souplesse pour satisfaire la plupart des besoins graphiques particuliers en terme d'échelle, de découpage, de projection et de sélection de l'information.

D'autre part, la création de base de données est une question très complexe qui nécessite des études théoriques et des analyses multidisciplinaires détaillées. La mise sur pied des bases de données a des répercussions budgétaires considérables et le système doit être efficace et viable à long terme. La création de la base de données géographiques du Québec déborde le cadre de la cartographie et fait actuellement l'objet d'une étude globale au sein de divers organismes gouvernementaux impliqués dans cette question.

DEROULEMENT DE LA PRODUCTION EN CARTOGRAPHIE DE BASE NUMERIQUE

En cartographie de base numérique la production est actuellement réalisée à deux échelles: 1:1 000 et 1:20 000. Même si la notion d'échelle a peu de signification lorsqu'on parle de données numériques, elle reste un critère très pratique pour caractériser la production.

Le premier projet de cartographie à grande échelle a débuté en septembre 1981 et concerne deux municipalités de l'Ile de Montréal: la ville de Montréal et la ville d'Outremont. La saisie de l'information est réalisée à partir de photographies à l'échelle 1:5 000. La production numérique pour ces municipalités s'inscrit dans le cadre d'ententes entre les municipalités et le ministère pour la création des bases géographiques urbaines. Les deux municipalités en question ont à leur disposition des systèmes graphiques interactifs. Le ministère fournit la carte topographique de base sous forme numérisée et les villes implantent tous les autres fichiers nécessaires à la gestion municipale (ingénierie, urbanisme, zonage, etc...).

La cartographie à l'échelle 1:20 000 a débuté en janvier 1983 dans la région du Témiscamingue, c'est-à-dire une région boisée, très peu développée et moyennement accidentée. Les éléments de la carte à l'échelle 1:20 000 seront partiellement intégrés à la base de données géographiques du Québec actuellement à l'étude.

La production est réalisée en trois étapes: le captage, l'édition et le traçage.

Le captage

Les fichiers de captage sont créés en prélevant directement l'information sur les photographies aériennes à partir des appareils de photogrammétrie. Tous les détails numérisés correspondent à des éléments graphiques qui sont automatiquement visualisés sur une station interactive. Les fichiers de captage contiennent la localisation précise des éléments avec un minimum de symbolisation.

Dans les fichiers de captage, l'information n'est pas codifiée puisque la production n'est pas orientée vers les bases de données. Cependant, les éléments graphiques sont structurés en utilisant les possibilités informatiques du système, ce qui permet une classification automatique des éléments de même nature. Chaque élément est déterminé par quatre paramètres:

- le niveau d'information: à l'échelle 1:1 000, on utilise environ une soixantaine de niveaux. A l'échelle 1:20 000 l'information est séparée sur deux fichiers, un fichier pour la planimétrie avec une soixantaine de niveaux et un fichier pour l'orographie avec une cinquantaine de niveaux utiles.
- le poids de l'élément qui correspond à la largeur de la ligne; une dizaine de poids différents sont utilisés, sur une possibilité de trente-deux.
- le type de ligne c'est-à-dire les lignes pleines ou tiretées: les quatre types de lignes du système sont utilisés
- le type d'élément: droite, courbe, forme géométrique, cellule, texte, etc...

A l'échelle 1:1 000, les fichiers de captage par modèle photogrammétrique sont directement utilisés pour l'édition cartographique. A l'échelle 1:20 000, une étape supplémentaire est ajoutée avec la création d'un fichier de captage par feuille cartographique et qui contient des corrections de tracé.

L'édition

Les fichiers d'édition sont produits par traitement des fichiers de captage, sur une station graphique interactive. Tous les éléments qui requièrent une symbolisation en vue d'une représentation graphique sont transformés et la toponymie est ajoutée.

Le traçage

Les fichiers de traçage traduisent les fichiers d'édition dans un langage directement assimilable par les traceurs automatiques. Le traçage est effectué sur table automatique de grande précision par méthode de gravure sur couche. Le produit final positif est réalisé de façon conventionnelle par reproduction photographique.

METHODOLOGIE ET SPECIFICATIONS EN CARTOGRAPHIE DE BASE NUMERIQUE

Depuis 1974, le ministère possède des normes en cartographie conventionnelle à grande échelle (1:1 000, 1:2 000, 1:2 500) et à moyenne échelle (1:5 000, 1:10 000, 1:20 000) autant pour la préparation des stéréominutes que pour le dessin des cartes. La préparation des normes en cartographie numérique a été une des premières étapes de la production. Les nouvelles normes se sont inspirées des normes conventionnelles et ont pris en considération les possibilités et les contraintes des systèmes informatiques. Actuellement, les normes pour la constitution des fichiers numériques à grande et à moyenne échelle sont disponibles à la Photocartothèque québécoise.

La production en cartographie numérique a nécessité le développement de nombreux logiciels d'application.

A l'étape du captage, la préparation d'un menu et de commandes spécifiques pour la saisie de l'information est essentielle à la rentabilité du projet en permettant de minimiser les manipulations informatiques par l'opérateur. Au cours des premiers mois de production, de nombreuses modifications ont été apportées à ces commandes, compte tenu des suggestions apportées par les opérateurs et les responsables de la production.

A l'étape de l'édition, il a été nécessaire de créer une librairie pour tous les symboles cartographiques et une librairie pour les lettrages esthétiques particuliers à la cartographie. A l'étape de l'édition, les développements informatiques sont beaucoup moins nombreux mais beaucoup plus complexes qu'à l'étape du captage car ils concernent l'automatisation de certaines opérations (jonction des segments d'éléments linéaires, symbolisation des lignes, etc...).

SUIVI DE LA PRODUCTION EN CARTOGRAPHIE DE BASE NUMERIQUE

Le ministère a gardé la même politique administrative pour la cartographie numérique que pour la cartographie conventionnelle c'est-à-dire que tous les travaux sont exécutés à contrat par l'entreprise privée. Cependant, la technologie étant très nouvelle et nécessitant des investissements importants, la forme du contrat de cartographie numérique est très différente des contrats de cartographie conventionnelle. Après de multiples rencontres avec toutes les entreprises privées du Québec, le ministère a octroyé un contrat de cinq ans à un Consortium composé d'une société d'état et d'une entreprise privée en cartographie. Ce contrat garanti permettait au Consortium d'acquérir l'équipement nécessaire à la nouvelle tehcnologie.

Le suivi de ce contrat exige une collaboration étroite entre le ministère et les contractants sur le plan technique et administratif.

Le rôle du ministère se situe à deux niveaux: le choix des critères techniques et le contrôle de la qualité, dans un souci constant de rentabilité.

Techniquement, le ministère avait conçu des méthodologies et des spécifications en tenant compte de l'expérience acquise dans les études préliminaires. Dans un contexte de production, l'entreprise privée a apporté de multiples suggestions pour améliorer les méthodes et diminuer les temps de production. Après des discussions avec le Consortium, le ministère évalue les recommandations et donne son accord à des changements de méthodes, au développement de logiciels ou à un assouplissement des normes. Les décisions doivent être prises rapidement et la collaboration entre les responsables de la production et ceux de la recherche doit être excellente.

La qualité de la production est suivie comme en cartographie conventionnelle. L'exactitude de l'interprétation des données photogrammétriques ainsi que le respect des normes sont vérifiés systématiquement sur un tracé rapide de chaque carte. Après les corrections demandées, un deuxième traçé rapide est envoyé au ministère ainsi qu'un ruban magnétique.

Le fichier numérique est analysé sur une station graphique, ce qui permet de compléter la vérification du produit graphique. Enfin, la précision photogrammétrique est vérifiée par échantillonnage en examinant certains modèles sur les stéréorestituteurs.

LE FICHIER DES LIMITES MUNICIPALES

Avec le fichier des limites municipales, le ministère aborde les applications de la technologie numérique dans le domaine des cartes thématiques.

Comme pour la carte photogrammétrique, le fichier des limites municipales est conçu pour des représentations graphiques mais permet de sélectionner certaines limites à partir de tris statistiques. Par conséquent, le fichier graphique est associé à un fichier alphanumérique et son exploitation requiert l'utilisation d'un logiciel de traitement de données.

L'information graphique est saisie en numérisant sur un digimètre les limites des municipalités préalablement tracées sur des cartes à l'échelle 1/20 000 ou 1/50 000. La numérisation a été réalisée par l'entreprise privée sur des systèmes interactifs ou aveugles selon des spécifications précises. Chaque carte numérisée constitue un fichier et le ministère a procédé aux étapes subséquentes:

- transformation des fichiers de chaque carte en coordonnées géographiques (latitude, longitude)
- Assemblage des fichiers par carte pour obtenir un fichier provincial avec l'ensemble des limites municipales.
- Création des polygones en assemblant les différents segments du contour de chaque municipalité.

Le fichier alphanumérique provient du Bureau de la statistique du Québec et contient un certain nombre d'informations reliées aux municipalités comme par exemple: le nom exact de la municipalité, un code à 5 chiffres, le nom de certaines entités administratives, etc...

Le numéro de la municipalité dans le fichier alphanumérique est associé à un centroïde pour chaque polygone du fichier graphique.

COMPARAISON ENTRE L'APPLICATION PHOTOGRAMMETRIQUE ET L'APPLICATION THEMATIQUE

Même si la technologie et les équipements utilisés en cargraphie numérique sont sensiblement les mêmes dans les deux types d'application, les problèmes rencontrés sont totalement différents.

La cartographie photogrammétrique est caractérisée par une grande diversité d'éléments graphiques constitués de points, de lignes et de surfaces. Les fichiers graphiques sont de très grandes dimensions et les développements informatiques sont importants pour la saisie de l'information. Le défi majeur réside dans l'exploitation optimale des possibilités graphiques des systèmes pour diminuer le temps d'opération.

En cartographie thématique, la préparation de l'information a une importance considérable. Toute l'information doit être compilée et vérifiée sur les documents graphiques avant tout traitement numérique. Dans les fichiers numériques, les éléments graphiques sont souvent des polygones et le traitement fait appel à des logiciels commerciaux complexes qui doivent souvent être adaptés à chaque situation.

CONCLUSION

Après deux ans de production, la cartographie de base numérique a atteint un rythme valable de productivité, compte tenu des équipements disponibles et des logiciels élaborés. La carte numérique 1/1000 rend possible la création des premiers systèmes numériques intégrés pour la gestion municipale. Avec les premières cartes numériques 1/20 000 et le fichier des limites municipales, le ministère possède un matériel de base pour expérimenter les recherches théoriques sur la base de données géographiques québécoise.

L'adaptation à la cartographie numérique constitue un défi pendant les premières années de production car cette technologie entraîne une foule de problèmes techniques mineurs totalement imprévus au moment de la conception administrative des projets. L'implantation de la technologie numérique exige autant de qualifications humaines que d'équipement technique. Un personnel scientifique polyvalent et innovateur soutenu par une administration flexible et rapide est un des atouts majeurs de la réussite en ce domaine.

AUTOMATED STANDARD NAUTICAL CHART PRODUCTION

Messrs. Larry Strewig, Joseph P. Ruys, Jan S. Schneier
Defense Mapping Agency
Washington, D.C. 20315

ABSTRACT

In September of 1982, the Defense Mapping Agency printed and released to the public a Standard Nautical Chart which was compiled, symbolized, and color separated using computer technology. The production of this chart involved several different computer driven mapping systems, as well as a unique mix of cartographers and computer specialists. This report describes the methodology, techniques, and computer systems used to produce the chart.

INTRODUCTION

In September 1982 the Defense Mapping Agency (DMA) released for sale to the public, chart 62097, Approaches to Bandar at Tawahi. The publication of this chart represented a milestone in the Defense Mapping Agency's continuing efforts to develop more efficient production procedures for its standard nautical charts.

The compilation of a standard nautical chart using conventional techniques normally involves a series of labor intensive, manual steps. Generally speaking, a cartographer will select information from various hydrographic and topographic sources for portrayal on a chart. This information is 'pulled-up' on a series of overlays, which are then mosaicked to fit a final scale, transformation, and datum, forming a compilation manuscript. The manuscript is then symbolized through scribing or drafting operations.

In the compilation of chart 62097, these same tasks were accomplished; but automated cartographic techniques and methodologies were employed, instead of the manual procedures outlined above. In addition, a set of constraints and objectives were established to ensure that a chart produced in this manner would be as accurate and cartographically correct as those compiled conventionally. These constraints and objectives are listed below.

CONSTRAINTS

The automated chart production system was to utilize only equipment currently in operation.

Product integrity and accuracy were not to be sacrificed for increased savings and throughput times.

OBJECTIVES

Compile a nautical chart utilizing automated means to the fullest extent possible.

Determine the optimal utilization of equipment and manpower.

Assess the advantages and disadvantages of the automated chart production system developed.

Minimize throughput times and maximize savings while maintaining product integrity.

The following discussion describes the techniques and methodologies, as well as the equipment, used to compile chart 62097.

CURRENT TECHNOLOGY AND COMPILATION REQUIREMENTS

An analysis of the source documents used in the compilation of a nautical chart would reveal that they basically fall into two categories: those used primarily for hydrographic information, and those used mainly for topographic data (see Figure 1. for chart 62097 sources). Closer analysis would reveal that the information selected from a hydrographic source for use in a chart is preponderantly point features, (soundings, aids to navigation, etc); while topographic information is primarily linear in nature. This distinction was a key factor in determining how to optimize the use of automated cartographic technology in the chart compilation procedure.

The automated cartographic equipment existing at the DMA today is either raster or vector in nature with regard to their data storage, manipulation, and capture capabilities. Each of these technologies have relative advantages and disadvantages in their use.

Raster scanning technology is extremely efficient in the mass digitization of source documents containing medium to dense line work. Using raster scanning methods, sources meeting these requirements can be digitized in a quick and efficient manner. This method of data capture does have one significant drawback however, the raster information cannot be assigned any attributable information (i.e., a bathymetric contour cannot be assigned a depth value). Without this information, the collected data would not be very useful in the compilation procedure, and could not be part of an integrated chart data base. Thus, to reap the benefits of mass digitization, the raster scanned data must be converted to vector mode; a data format which permits the assignment of descriptive information to digitized data. The cartographer, by manipulating data in this fashion avoids the tedious and time consuming chore of manually collecting this data with a hand held digitizer.

As mentioned earlier, data in vector format has the advantage of permitting 'tagging'. This format is also advantageous because it requires less storage space than the raster format, and is more efficiently manipulated especially with regard to compilation functions. Such operations as projection transformation, scale change, mosaicking, merging, sounding unit conversions and general distortion removal are more easily and efficiently accomplished with data in vector format.

One additional capability must be discussed regarding vector based technology; the capture, tagging, and processing of point feature data. To incorporate a point feature into a vector file, a cartographer must build a header describing that feature, 'visit' its location, and digitize its position and attribute information. While this procedure may seem relatively inefficient especially with regard to the mass digitization of linear features, there is currently no existing 'automated' alternative at DMA within the nautical chart program.

Given the relative merits of each of the two technologies, as well as the differences between hydrographic and topographic feature types the decision to maximize the use of DMA's automated cartographic equipment in the chart compilation process was clear cut. A raster based system, the Sci-Tex Response 250 Mapping System, was utilized in the mass digitization/vectorization of topographic sources. The Advanced Cartographic Data Digitizing System (ACDDS), a vector system, was used to collect the hydrographic data and perform chart compilation tasks.

EQUIPMENT DESCRIPTIONS

ADVANCED CARTOGRAPHIC DATA DIGITIZING SYSTEM (ACDDS)

The ACDDS is a vector formatted digitization and compilation system. It consists of a host subsystem and four (4) independent workstations.

The ACDDS was developed for DMA by the Synectics Corporation. The system hardware was designed to be redundant so if any part of the system fails, the rest of the system can continue operation.

System Procedures - Analog data, including survey sheets, maps, U.S. charts, foreign charts and other manuscripts, are digitized on the digitizing table and stored on the workstations disk at their original scale and projection. The data are digitized in one of three modes: trace, depth (soundings), or point feature. As each feature is digitized, it is displayed on the graphics CRT for verification by the cartographer. At any point in this digitizing process the cartographer has available a full range of interactive edit functions to modify any feature in the file.

When this data is correct the cartographer can transfer the data to the host system to perform any of the following batch functions:

Projection transformations including datum shifts and scale changes.

Sectioning - data inside a cartographer-defined area are extracted and put into a separate file.

Merging and paneling - several files are merged into one file.

Sorting (filtering) - feature types defined by the cartographer are extracted and put into another file.

Unit conversion - the units of measure of soundings are converted to other units of measure (i.e. fathoms and feet to meters).

Proof plotting - either symbolized or centerline data depiction.

Magnetic tape input and output.

The information collected from various sources at the edit station are combined into one master file of the same scale and projection type through host operations. This file can then be returned to the edit station for additional interactive edits and refinements. Such edits include the addition of any new data (ie., Notice to Mariners), the correction of any overprinting information, and the improvement of the aesthetic qualities of the chart.

When the data is fully edited and in final form the cartographer can send the file to the host system where fully symbolized and color separated CRT plotter tapes can be generated. These tapes will drive the plotter to produce reproducible quality film positives.

THE CRT PRINTHEAD PLOTTER

The CRT Printhead is a very high speed, vector formatted plotter capable of producing reproducible quality film positives. This system was developed for DMA by Image Graphic Inc. It allows dense vector formatted data to be plotted at a speed equal to or greater than raster plotters. Data are sorted into 2" x 2" pages before plotting to minimize the mechanical movement of the plotter. When the CRT printhead plotter receives this sorted data, it moves the printhead to each "page" and flashes all of the data within that 2" x 2" area. This procedure greatly reduces the time consuming mechanical movement characteristic of conventional vector plotters.

The CRT plotter can also virtually store an unlimited number of reproducible quality symbols on its software disks, thereby making point symbol and type placement economical.

THE SCI-TEX RESPONSE 250 MAPPING SYSTEM

The SCI-TEX Response 250 Mapping System (R250 MS) is a color, raster/graphic editing system with a scanner, edit station, and film exposure device (a laser plotter). The flow of data through the system is from scanner (data capture) to edit station (color separation corrections and cartographic updates), then to laser plotter (color separated film output).

The SCI-TEX Response 250 Mapping System was acquired "off the shelf" in late 1978/early 1979. The configuration of the system at DMAHTC currently consists of one scanner, 4 edit stations, and one film exposing device (a laser plotter).

System Procedures- Analog data including survey sheets, maps, U.S. charts, foreign charts and other manuscripts are mass digitized via a 12 color raster scanner. Data are stored on edit station disks as functions of x, y position and color. These data can be scanned at various resolutions (4 - 47 points per millimeter) accessed and displayed at the edit station.

Depending upon the nature of the final product (i.e., color separation, resymbolization, etc), the following edits may be utilized:

Raster/Vector/Raster data conversions - Linear data captured via the scanner in raster mode can be converted to vector mode, tagged with a symbolized feature type such as a built-up area boundary and converted back to raster as fully symbolized data.

Point feature symbolization - Point feature data can be symbolized by calling various symbolized point features from a previously defined library and inserting them into a file at a desired x, y location, at any rotation. The insertion of symbols into the raster data can be accomplished interactively, manually, or automatically.

Area Feature Symbolization - Area outlines can be defined interactively or acquired via the scanner. The areas defined by these outlines can be filled with a solid color (thereby producing an open window of the area) or a symbol can be repeated throughout the area (i.e. a marshland symbol could be repeated across a polygon producing a marshland areal symbol).

Color Separation Editing - Scanned data inevitably contains data which has been incorrectly sensed and color coded (i.e. dark blue data sensed and coded as a light blue). These areas or islands of incorrectly color coded data can be recoded to their correct color via interactive or global (file wide) editing functions.

Type Composition and Placement - Typed in various fonts and point sizes can be composed and interactively placed into a data file. Currently the DMAHTC R2500MS font Library consists of the following type styles and point sizes:

News Gothic Condensed - 6 and 7 point
News Gothic - 6,7,8,10,12,14,16,18 and 30 point
Techno Medium - 6,7,8,10,12 and 14 point
Techno Medium Italic - 6,7,8,10,12 and 14 point
Century Expanded - 6,7,8,10,12 and 14 point
Century Expanded Italic - 6,7,8,10,12 and 14 point

When the data are fully edited and in final form the cartographer can output his data file as color separated film negatives or positives via the laser plotter.

COMPILATION PROCEDURES

TOPOGRAPHY (see figure 2)

The topographic sources used in the compilation of Chart 62097 were photogrammetrically derived manuscripts, compiled on the AS11A stereo compilation instrument. The AS11A through interior, relative, and absolute-orientation of stereo imagery was used to remove film and modal distortions (tilt, parallex, etc.) from the photographic coverage of Aden Harbor. The photogrammetrist then utilized this stereo imagery to extract significant topographic features. These data were compiled at the final scale and projection of the chart. Additional features necessary for hydrographic/navigation requirements were then added to these manuscripts through a landmass intensification procedure.

The linear data depicted on any topographic source can be classified in several ways. These data can be classead according to the information they represent (i.e., a road vs. a railroad), or the manner in which they are depicted on the document (double line symbol (road) vs. a ticked line (railroad)). To facilitate the mass digitization and subsequent symbolization of the topographic manuscript, the compiled linear features were classed by the latter method. This classification was accomplished through the use of several different colors as data depiction mediums.

For example those features that were to be symbolized (on the final chart) as solid lines with a lineweight of 8 mils (shoreline and drainage) were depicted as centerline information in the color red, while those requiring cased or paralleled line symbolization were drawn as centerline information in the color green. The net result of these actions was a symbol classification manuscript that facilitated the symbolization of topographic information.

The manuscript was mass digitized (scanned) on the Sci-Tex scanner at a resolution of 10 points per millimeter (x0.004 mils). The scan required 7 colors, 6 for the color coded information, and one channel for the background color. Upon completion of the scan, the resulting data (in raster format) were checked and corrected for minor dimensional inaccuracies. Following this procedure, the scanned information was examined (at the edit station) for any improperly color coded data induced by the scanner.

As a necessary step to achieve the conversion of the data to vector format, a skeletonization operation was initiated. This function automatically delineated the center or midline of each feature. These skeletonized data were then edited for unintended gaps and spurious data (i.e., spikes generated by skeletonization). With these edits complete, the topographic information was ready for vectorization.

The program which initiates the conversion of skeletonized data to vector strings requires that the operator assign a symbol tag to the information to be vectorized. It is at this juncture that the color coding of topographic linework comes into play. By guaranteeing that those features coded in a given color are to be symbolized identically on the chart, we can automatically tag each feature within a classification with the same symbol tag. So when the vector data is rasterized again, the correct reproduction quality symbology has replaced the original pencil compilation.

The photogrammetric manuscript contained a limited number of point features which were interactively symbolized via the Sci-Tex. By utilizing a predefined symbol file and visiting each point feature, the appropriate symbol was incorporated into the data (i.e. a rock awash or building symbol).

HYDROGRAPHY (See Figure 3)

The film positive generated on the Sci-Tex served as the controlling source for the compilation of hydrography on the ACDDS. All 5 hydrographic sources used in the compilation were controlled through shoreline matches to the Sci-Tex, topographic positive.

Each piece of hydrographic source material was registered to an ACDDS digitizing table and features selected by the cartographer were digitized at the scale, unit of measure and projection of the source document. This digitization was done by selecting a header for each feature, entering a depth or elevation if necessary, and manually digitizing each feature with a cursor. As these data were digitized they were displayed on the system's graphic CRT to allow the cartographer to verify these data and view their spatial relationship to other information. The digital file of each source was then plotted on the Xynetics proof plotter. This plot was then compared with the original source document to verify the accuracy and completeness of the digital file. If data in any of the source files were found to be incorrect or missing the file was then corrected interactively by the cartographer on the ACDDS workstation.

Some of the sources used for chart 62097 had soundings measured in fathoms and feet rather than meters. Since U.S. Chart specifications require all soundings to be shown in meters, the sounding features of these sources were converted to meters automatically by the ACDDS. This eliminated the labor intensive effort of using a table to look up the meter value of each sounding. It also eliminated the possibility of human error during the conversion process.

Following these operations, the digital file of each source was mathematically transformed by the ACDDS to the projection, scale and spheroid of the final product, and then digitally mosaicked to the shoreline of the Sci-Tex symbolized positive and/or other hydrography. This was accomplished by sectioning each source file, shifting the data of that section to fit the controlling source and then merging all of these sections into one master file.

After all of the sources were merged into one file, a proof plot was generated to allow the cartographer to determine what corrections, additions, and/or modifications had to be made. The data file was then interactively modified at an ACDDS workstation on a feature by feature basis. This modification process may include:

Joining (connecting) lines
Smoothing lines
Modifying segments to avoid overprints of data
Deleting unnecessary data
Adding newly acquired data (i.e. Notice to Mariners corrections)
Other feature modifications necessary to make the final product more useful and aesthetically pleasing.

When the cartographer was satisfied with his completed compilation, a CRT Printhead plot tape was generated. Since each feature was tagged with attribute information as it was digitized, the plot tape generated contained the necessary instructions to direct the CRT Plotter in symbolizing each feature. The actual film positive production time on the CRT was significantly less than with conventional line following plotters, due to the exposure of 2" by 2" windows of data at the same time.

TYPE GENERATION AND PLACEMENT

Following the completion of the hydrographic compilation phase of the production chart 62097, a type guide was prepared upon which was depicted the approximate location for a string of text, its point size and font type. This guide (prepared on a proof plot of the hydrographic data) was then scanned on the Sci-Tex and merged with the previously symbolized topography. This file served as a reference pattern relative to which strings of text could be interactively placed on the Sci-Tex Edit Stations.

At the system's prompting, the cartographer was requested to key in the text needed (up to 99 words in a pass). These strings were then automatically composed and placed into a file that was accessed for interactive placement. Utilizing the data file of the merged type guide/topography as a reference pattern and a duplicate file of the previously generated symbolized topography as an output file, the cartographer proceeded to interactively place the text. When deemed necessary, a paper verification plot was generated for edit purposes. The cartographer would note on this plot any missing text, misspelled words, lines that needed to be broken, or text that needed to be placed in a new position. When all the corrections were completed, final color separated film positives of the text and topography were generated, and a data tape generated for future use.

OPEN WINDOW AND AREA SYMBOLIZATION

A Xynetics plot was generated on the ACDDS of bathymetric contour data, coral, dangers, and uncovering area. A color dot was placed into each of the areas requiring open window symbolization; one color for areas of coral and uncovers, and a different color for areas of danger and significant depth (from the shoreline to the 5 meter curve). Where these curves were broken (interrupted) they were closed to ensure that each color dot was surrounded by an uninterrupted polygon. This tint guide was scanned and merged to the existing symbolized topographic data (the land tint, built up area, and saltpans were generated previously during the processing of the manuscript). These data were then checked for accuracy.

Utilizing a color spreading algorithm, the areas enclosed by each polygon were painted 'solid' using the color dot as the active color. The bounding polygon was then deleted, being replaced by the area fill color(s). In this manner islands of color were created that, when viewed as a integral unit, formed the open window tint area symbolization. A verification paper plot of this data was generated and adjustments made where necessary. This edited file was then used to generate open window film negative via the plotter and the data were saved for future use.

The reproduction quality film positives and negatives generated by the CRT Printhead and Sci-Tex plotters were photo composed and processed in the conventional fashion, concluding the compilation and production of chart 62097.

BENEFITS OF THE AUTOMATED CHART PRODUCTION SYSTEM

The ultimate success of any new production procedure is determined by how it compares to the previous system in terms of accuracy, total manhours and associated pipeline days. In the case of the system just described, significant savings in both manhours and pipeline days will be realized in chart production as a result of the system's initial design and future refinements. In addition the accuracy of the resulting chart was equal to or better than those compiled conventionally.

Utilizing this system, a savings of approximately 26% in manhours can be expected for an average chart. This savings will result from the elimination of many tasks previously required during compilation, (i.e. automatic sounding unit conversions), as well as decrease in the number of redundant operations performed by the cartographer.

The production of a chart through conventional procedures normally requires a good deal of support work from branches other than the compiling element. Utilizing the procedures discussed, most of the support activities normally associated with conventional production procedures (type-setting, drafting/scribing, negative engraving, and photographic processing) are either eliminated altogether, or greatly reduced in time. The pipeline time resulting from the movement of the job from work center to work center to perform these support activities will be reduced significantly (approximately 35%). An example is the virtual elimination of the 100 day scribe/draft production step throughout the use of symbolization algorithms and finishing plotters to produce the reproducible quality color separates.

In addition to the economic benefits resulting from the initial compilation of the chart, further savings will accrue as second and third editions of the chart are required. Subsequent to the publication of a nautical chart, a record is kept of those changes occuring in the area of the chart. When the number of the changes exceeds a certain limit, a new edition of that chart is called for. By maintaining the data resulting from the initial compilation of a chart in a digital chart library, new editions of these charts can be easily produced. Utilizing the initial digital data, the required changes can be made through the system and a new set of reproduction quality positives generated. Any subsequent editions of the chart will be produced in a more efficient manner than could otherwise be accomplished; thus leading to additional savings.

SYSTEM REFINEMENTS

During the compilation of 62097, there were certain inefficiencies in the procedures identified. These inefficiencies, specifically, type placement and open window negative generation, were significant enough to warrant a re-examination and correction of the procedures utilized. While they were not remedied soon enough to be useful on the compilation of chart 62097, the corrections are being applied to current production.

The type placement, while accomplished successfully on the Sci-Tex, proved to be a lengthy process in terms of the time needed to prepare the type guide and the incorporation of the type into the file. To correct this problem, a type placement system has been developed for use on the ACDDS. The cartographer using this new software can efficiently position type strings without the need of a type guide, as well as determine the impact a particular string of text has on other data in that area. The actual 'flashing' of the type is done on the CRT Printhead Plotter, avoiding the 'update' time inherent on the Sci-Tex system.

The open window negative generation scenerio was basically a cumbersome procedure, and for this reason not as efficient as could be possible. Under this initial procedure, the cartographer was required to manually create a tint guide. The solution devised for this problem was to create a data link from the ACDDS to the Sci-Tex. This will permit the cartographer to transfer his polygon data digitally (tape transfer) to the Sci-Tex, where the painting algorithm is used as described earlier.

Finally, a third improvement was made to the system. Software was developed that permits the transfer of vectorized topographic data to the ACDDS from the Sci-Tex (the reverse of the other data link). This transfer will, among other things, permit the entire data file resulting from the production of the chart to be stored in one file.

CONCLUSION

The publication of standard nautical chart 62097 represents a significant step forward for DMA's hydrographic automated cartography effort. The production of this chart realized the initial objectives of the project without compromising any constraints. As a result, the automated cartographic production of standard nautical charts can be viewed as coming of age at the Defense Mapping Agency.

A STEREO ELECTRO-OPTICAL LINE IMAGER FOR AUTOMATED MAPPING

J. R. Gibson, R. A. O'Neil, R. A. Neville, S. M. Till, W. D. McColl
Canada Centre for Remote Sensing
2464 Sheffield Road, Ottawa, Ontario
Canada K1A 0Y7

ABSTRACT

This paper discusses the rationale for building a demonstration single pass stereo line imaging system and provides some of the background developments that have made such a project possible at this time. The paper discusses the integration of the postion and attitude data from an inertial navigation system into the absolute orientation of the imagery data. Then the procedure for the geometric correction of the imagery is shown in detail. Finally a line imager based stereo mapping system is proposed.

INTRODUCTION

CCRS is currently embarking on a program to enhance an existing electro-optical line imager to allow the acquisition of fore/aft stereo data (single pass stereo) as well as to put into place a system to compute the geometric correction for the imagery using simultaneously acquired navigation data. This paper describes in detail the stereo line imager and the geometric correction process and also outlines a potential end-to-end digital mapping system.

The demand for airborne line imager data for precision work such as mapping has not been great in the past. There have been two fundamental reasons for this; the pixel resolution has typically been very coarse and the spatial distortions due to uncompensated aircraft motion and attitude have degraded the imagery.

Improvements in technology have been making it possible to think in terms of overcoming these drawbacks to line imager data. At the same time, changes in the approach to mapping are occurring which makes a direct digital input of data more attractive than has been the case in the past.

The challenge for the line imager technology is clear: Produce system for digital mapping. The advantages of a digitally based mapping system make it worthwhile to take up the challenge. The opportunities to achieve improvements in quality and throughput lie in the following attributes of an electro-optical line imager and associated computer-based processing system:

1. Currently available electro-optical elements have resolving capabilities that are approaching the grain size of photographic emulsions. The stigma of poor resolution will soon pass.

2. The increased dynamic range of the electro-optical elements vs film will provide more detail and greater radiometric accuracy for subsequent processing e.g. automatic correlation for elevation data.

3. By displaying the stereo data on CRT, the expensive and restrictive optical trains of existing plotters may be eliminated. Digital processing for rotations, enlarging and image enhancements will inherently be more flexible and versatile than the fixed operations available from the optics of a plotter.

4. Since each pixel will have accurately computed X and Y coordinates, the need for the expensive electro-mechanical measuring components in current plotters will be eliminated along with their associated problems.

5. The geometric fidelity of a line imager is potentially much greater than that of film. Once an imager has been calibrated (subject to periodic verification), the problems of film induced distortions will be eliminated.

6. A stereo line imager will provide continuous strips of stereo data as opposed to the overlapping segmented data sets provided by photographs. This will eliminate the non-trival task of linking segmented data sets together and allow for longer unbroken runs of data through the elevation determination programs.

CCRS PUSHBROOM IMAGER

The Canada Centre for Remote Sensing has developed a second generation pushbroom line imager, MEIS II. This imaging sensor uses silicon charge-coupled device linear array detectors with usable response covering the 380-1100 nm spectral range in the visible and near infrared. The sensor is characterized by high spatial resolution and low noise equivalent radiance making it particularly attractive for applications in resource classification, vegetation stress analysis, passive bathymetry, digital mapping and digital elevation modelling.

MEIS II is an airborne multispectral imaging sensor of the "pushbroom scanner" type. This type of imager utilizes multi-element linear array detectors oriented perpendicular to the flight direction and located in the focal plane of the sensor optics as illustrated in

Figure 1. Each detector in the array accumulates a charge which is proportional to the light energy incident upon that element during the time interval between read-outs. These charges are electronically sampled to produce a line image of the scene below. The second dimension in the image is provided by the forward motion of the aircraft. That is, with the detectors used in MEIS II, the sampling of the photosites is virtually simultaneous for all elements, within the contraints imposed by the finite pulse propagation speed. In the normal mode of operation, the forward motion during each integration period is equal to the lateral distance mapped onto an individual array element.

The pushbroom imager enjoys an advantage in terms of sensitivity and spatial precision, as compared to mechanically scanned imagers. The increased sensitivity stems from the much longer integration period available for each pixel. Whereas, in the mechanical scanner a single detector is swept across the target, sampled once for each pixel, to build up a line image, the pushbroom imager devotes one detector to each pixel for the whole duration between image line readouts. This gives an integration period for the pushbroom imager that is larger by at least a factor equal to the number of elements in an image line. This increased radiometric sensitivity permits imaging under lower light conditions or over darker targets. The increased spatial precision of the pushbroom imager results from the fact that the optics and detector array geometry is fixed, and therefore can be precisely calibrated in the laboratory. The mechanical scanner requires the precise dynamical measurement and control of its moving component, a much more demanding task. The improved spatial precision and geometric fidelity of the pushbroom imager are features required in cartography and terrain modelling. Other advantages offered by the pushbroom imager include a simpler mechanical and optical design and a greater operating flexibility. In the present design, spectral bands are readily selected by the appropriate choice of front-mounted filters. The angular field of view can be changed by the replacement of the lenses by ones of different focal lengths.

The MEIS II imager has provided an ideal base for demonstrating a stereo imager system. The separate optical channel configuration has enabled the attachment of external mirrors to create a forward looking and an aft looking channel in addition to the normal downward looking channels. This is shown schematically in Figure 2. The mirror mount is currently being fabricated and test flights will take place shortly.

Stereo line imagers have been proposed in the past, however, highly accurate position and attitude data to correct the imagery has always been recognized as being an essential component for such devices and the instruments required to provide this data have not been available (Masry, Derenyi). Over the past several years CCRS has developed the capability to acquire and process data from an LTN-51 Inertial Navigation System (INS), and in so doing we have come to know the error characteristics of that system and the limitations of its performance. The typical worst case error figures for the LTN-51 are as follows: There is a linear drift in position with a rate of the order of 0.25 m/s to which is added a 0.5 m/s sinusoidal velocity error with a period of 84.4 minutes. The latter is known as the Schuler oscillation. The attitude errors of the system due to the position or velocity errors are typically less than 20 arc-seconds (0.1 milliradians) however to that are added errors due to resolver non-linearities and friction induced motion of the platform. The total attitude errors are thus of the order of two arc minutes (0.6 milliradians). There is the potential to reduce the angle errors by the use of newer technology in inertial systems. Once the low frequency position errors have been removed from the inertial data, the relative position errors have been found to be of the order of 1 to 2 m RMS. CCRS has also developed a high altitude laser profiler in cooperation with Optech Ltd. in Toronto. This unit has been successfully flown at a height of 10 km over mountainous terrain in the Kananaskis Valley, south west of Calgary, where the elevation changes are of the order of 1500 m. The data set has not been fully evaluated yet but appears to be very good.

STEREO IMAGER DATA PROCESSING

In processing the stereo imagery data, it is necessary to follow the route of obtaining the relative and absolute orientation parameters as is the case with conventional photography. In this case, the process is somewhat more complex because of the requirement to remove the low frequency errors of the inertial system. Figure 3 illustrates the geometry of the problem, and the following definitions apply to the vectors shown in the drawing:

- flight position $\underline{p}_m^T(t) = (x\ (t), y(t), z(t))$
- flight attitude $\underline{a}^T(t) = (r_m(t), p(t), h(t))$
- ground point coordinates $\underline{g}^T = (xg, yg, zg)$
- imager position and attitude at time t_k : $\underline{p}_k$, $\underline{a}_K$
- imager position and attitude at time t_n : $\underline{p}_n$, $\underline{a}_n$, that is $\underline{p}_k = \underline{p}\ (t_k)$, etc.
- pixel coordinates in imager frame: $\underline{m}_k$, $\underline{m}_n$
- pixel coordinates converted to Reference Frame; $\underline{x}_k$, $\underline{x}_n$

where $\underline{x}_k = C_k\ \underline{m}_k$ (1a)

and $\underline{x}_n = C_n\ \underline{m}_n$ (1b) and C_k and C_n are direction cosine matrices (rotation matrices) constructed from the attitude vectors $\underline{a}_k$ and $\underline{a}_n$.

- the base vector between the positions $\underline{p}_k$ and $\underline{p}_n$:

$$\underline{b} = \underline{p}_n - \underline{p}_k \qquad (2)$$

For an unknown ground point g, we have from the coplanarity condition.

$$\underline{b} \cdot \underline{x}_k \times \underline{x}_n = 0$$

or $$\underline{b} \cdot \underline{d} = 0 \qquad (3)$$

where $$\underline{d} = \underline{x}_k \times \underline{x}_n \qquad (4)$$

For ground control points we have:

$$\underline{p}_k + s_k \underline{x}_k = \underline{g} \qquad (5)$$

or $$s_k \underline{x}_k = \underline{v}_k \qquad (6)$$

where $$\underline{v}_k = \underline{g} - \underline{p}_k \qquad (7)$$

The scale factor s_k may be eliminated because Equation (6) is an expression of the fact that $\underline{x}_k$ is parallel to $\underline{v}_k$ so that their vector cross product is zero. Thus for control points,

$$\underline{x}_k \times \underline{v}_K = \underline{0} \qquad (8)$$

where $\underline{0}$ is a zero vector

Equations (3) and (8) may then be used as the basis for a least squares solution to solve for the low frequency position and attitude errors of the INS data. The operating assumption is that over short periods of time (10 minutes for example), the data supplied from the INS will have additive errors consisting of constant offsets plus linear drifts. That is we will be solving for corrections in position $d\underline{p}$ and corrections in attitude $d\underline{a}$ where these terms are assumed to have the form:

$$d\underline{p} = d\underline{p}_0 + d\underline{p}_1 t \qquad (9)$$

and $$d\underline{a} = d\underline{a}_0 + d\underline{a}_1 t \qquad (10)$$

The complete derivation for the least squares formulation may be found in Appendix A. The results are as follows, where it should be noted that upper case letters corresponding to vectors represent the skew-symmetric matrix that may be generated from the vector. For example, if:

$$\underline{x} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}, \text{ then } X = \begin{pmatrix} 0 & c & -b \\ -c & 0 & a \\ b & -a & 0 \end{pmatrix} .$$

The results of linearizing Equation (3) are:

$$\left(\underline{0}^T \quad -\underline{d}^T(t_n - t_k) \quad \underline{b}^T(X_k X_n - X_n X_k) \quad \underline{b}^T(X_k X_n t_n - X_n X_k t_k)\right) \begin{pmatrix} d\underline{p}_o \\ d\underline{p}_1 \\ d\underline{a}_o \\ d\underline{a}_1 \end{pmatrix} = (\underline{b}^T \ \underline{d}_o) \qquad (11)$$

Then from linearizing Equation (8), we have:

$$(X_k \quad X_k t_k \quad V_k X_k \quad V_k X_k t_k) \begin{pmatrix} d\underline{p}_o \\ d\underline{p}_1 \\ d\underline{a}_o \\ d\underline{a}_1 \end{pmatrix} = (\underline{x}_k \times \underline{v}_o) \qquad (12)$$

Note that on the right hand side of Equations (11) and (12), that the subscript "o" refers to the current estimate of the designated quantity. Thus, Equations (11) and (12) may be used to solve for the INS errors in an iterative least squares procedure. That is, both (11) and (12) may be expressed in the general form $A_i \, d\underline{u} = \underline{b}_i$. When a sufficient set of control and unknown points have been processed and accumulated in an A matrix and a $\underline{b}$ vector where:

$$A^T = (A_1^T, A_2^T, \ldots, A_n^T) \qquad (13)$$

and

$$\underline{b}^T = (\underline{b}_1^T, \underline{b}_2^T, \ldots \underline{b}_n^T) \qquad (14)$$

the least squares solution may proceed in the conventional manner as:

$$d\underline{u} = (A^T A)^{-1} A^T \underline{b} \qquad (15)$$

where $d\underline{u}^T = (d_{\underline{p}o}^T, d_{\underline{p}1}^T, d_{\underline{a}o}^T, d_{\underline{a}1}^T)$

GEOMETRIC CORRECTION

Once the INS data has been corrected, the imagery may be geometrically corrected in preparation for further processing. This is a relatively simple process consisting of recomputing the pixel coordinates to their local level equivalents. That is, during flight, the imager coordinate frame moves about as the aircraft changes attitude. Since we are able to measure the aircraft attitude we may compensate for the attitude changes as follows: The direction cosine matrix which is generated from roll, pitch and heading data defines for each image line, the rotation of the imager axes with respect to local level. This matrix may thus be used to convert the rotated pixel coordinates back to local level in the following manner. Figure 4 is a representation of the imager geometry. Each pixel represents a vector in the imager coordinate frame. That is:

$$\underline{r}_s = \begin{pmatrix} x \\ y \\ f \end{pmatrix} \tag{16}$$

where x is $\pm x_0$, the stereo offset distance
y is the "measurement" associated with each pixel
and f is the imager focal length.

As mentioned, the vector $\underline{r}_s$ is in the imager frame and it is necessary to convert this to the equivalent vector $\underline{r}_l$ in the local level frame. This is accomplished as follows:

$$\underline{r}_l = C^T \underline{r}_s \tag{17}$$

Expanding Equation (17) to show the elements of the vector gives:

$$\underline{r}_l = \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} C_{11} & C_{21} & C_{31} \\ C_{12} & C_{22} & C_{32} \\ C_{13} & C_{23} & C_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ f \end{pmatrix} \tag{18}$$

where it should be noted now that $z' \neq f$. Thus it is necessary to rescale each vector $\underline{r}_l$ in the local level frome to obtain the intersection of the vector with the local level plane. This is accomplished by multiplying the elements of each vector $\underline{r}_l$ by f/z' to give:

$$\frac{f}{z'} \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} x'' \\ y'' \\ f \end{pmatrix} \tag{19}$$

In this fashion, every scan line may be recomputed to its local level equivalent at which time it is possible to compensate for the effects of velocity variations and aircraft elevation changes.

STEREO IMAGER PROCESSING SYSTEM

The above procedures complete the data preparation stage. The data set is now ready for further cartographic operations such as DTM generation, contouring or plotting. A discussion of these operations is beyond the scope of this paper, however in summary, most of the technology appears to be available to put together a system to demonstrate the capabilities of a stereo line imager as a data source for an automated cartographic system. A block diagram of a possible system implementation is shown in Figure 5.

APPENDIX A

This section describes the generation of the least squares equations for solving for the INS error terms. Since this development is carried out using vector/matrix notation, the following vector derivative functions are used:

$$\frac{\partial}{\partial \underline{a}} (\underline{a} \cdot \underline{b}) = \underline{b} \qquad \text{(Gelb, p.22)} \qquad (A.1)$$

$$\frac{\partial}{\partial \underline{b}} (\underline{a} \cdot \underline{b}) = \underline{a} \qquad (A.2)$$

$$\frac{\partial}{\partial \underline{a}} (\underline{a} \times \underline{b})^T = -B \qquad (A.3)$$

$$\frac{\partial}{\partial \underline{b}} (\underline{a} \times \underline{b})^T = A \qquad (A.4)$$

where, as before, an upper case letter represents the skew-symmetric matrix form of a vector.

The derivative of a rotated vector with respect to the rotation angles is given by the following:

Let $\underline{x} = C\underline{m}$, (A.5),

then $d\underline{x} = d(C\underline{m})$

$= dC\underline{m} + C{\cdot}d\underline{m}$ (A.6),

where $dC = dAC$ (Britting, p.27) (A.7),

and dA is the skew-symmetric form of the angle derivative vector $d\underline{a}$.

Assuming that the vector $d\underline{m} = \underline{0}$, we have from Equation (A.6)

$d\underline{x} = dAC\underline{m}$

or $d\underline{x} = dA\underline{x}$ (A.8)

Equation (A.8) may be expressed as a cross product of $d\underline{a}$ and $\underline{x}$, giving:

$d\underline{x} = \underline{x} \times d\underline{a}$ (A.9),

and then using Equation (A.4) we have:

$$\frac{\partial \underline{x}^T}{\partial \underline{a}} = X \qquad (A.10).$$

Now turning to Equation (3) in the paper, this equation may be linearized as follows:

From, $f = \underline{b} \cdot \underline{x}_k \times \underline{x}_n$ (A.11),

and using a single term Taylor series expansion we have:

$$f = f_o + \left(\frac{\partial f^T}{\partial \underline{p}_k}\right) d\underline{p}_k + \left(\frac{\partial f^T}{\partial \underline{a}_k}\right) d\underline{a}_k + \left(\frac{\partial f^T}{\partial \underline{p}_n}\right) d\underline{p}_n + \left(\frac{\partial f^T}{\partial \underline{a}_n}\right) d\underline{a}_n \quad (A.12)$$

Since neither $\underline{p}$ nor $\underline{a}$ appear explicitly in (A.12) we must use the chain rule for differentiation to give:

$$\frac{\partial f^T}{\partial \underline{p}_k} = \left(\frac{\partial f^T}{\partial \underline{b}}\right)\left(\frac{\partial \underline{b}^T}{\partial \underline{p}_k}\right) = (\underline{x}_k \times \underline{x}_n)^T (-I)$$
$$= -\underline{d}^T \quad (A.13)$$

and similarly:

$$\frac{\partial f^T}{\partial \underline{p}_n} = \underline{d}^T \quad (A.14),$$

$$\frac{\partial f^T}{\partial \underline{a}_k} = \left(\frac{\partial f^T}{\partial \underline{x}_k}\right)\left(\frac{\partial \underline{x}_k^T}{\partial \underline{a}_k}\right) = (\underline{x}_n \times \underline{b})^T X_k$$
$$= \underline{b}^T X_n X_k \quad (A.15)$$

and
$$\frac{\partial f^T}{\partial \underline{a}_n} = \left(\frac{\partial f^T}{\partial \underline{x}_n}\right)\left(\frac{\partial \underline{x}_n^T}{\partial \underline{a}_n}\right) = -\underline{b}^T X_k X_n \quad (A.16)$$

Substituting Equations (A.13-16) back into (A.12) gives:

$$(-\underline{d}^T \quad \underline{b}^T X_n X_k \quad \underline{d}^T \quad -\underline{b}^T X_k X_n) \begin{pmatrix} d\underline{p}_k \\ d\underline{a}_k \\ d\underline{p}_n \\ d\underline{a}_n \end{pmatrix} = (-f_o) \quad (A.17)$$

Since
$$d\underline{p}_k = d\underline{p}_o + d\underline{p}_1 \; t_k \quad (A.18)$$
$$d\underline{a}_k = d\underline{a}_o + d\underline{a}_1 \; t_k \quad (A.19)$$
$$d\underline{p}_n = d\underline{p}_o + d\underline{p}_1 \; t_n \quad (A.20)$$

and
$$d\underline{a}_n = d\underline{a}_o + d\underline{a}_1 \; t_n \quad (A.21)$$

Equation (A.17) may be rewritten to give:

$$(\underline{0}^T \quad -\underline{d}^T(t_n - t_k) \quad \underline{b}^T(X_k X_n - X_n X_k) \quad \underline{b}^T(X_k X_n t_n - X_n X_k t_k)) \begin{pmatrix} d\underline{p}_o \\ d\underline{p}_1 \\ d\underline{a}_o \\ d\underline{a}_1 \end{pmatrix}$$
$$= (f_o) \quad (A.22)$$

Then looking at Equation (8) for the control points, we have:

$$\underline{f} = \underline{x}_k \times \underline{v}_k = \underline{0} \quad (A.23),$$

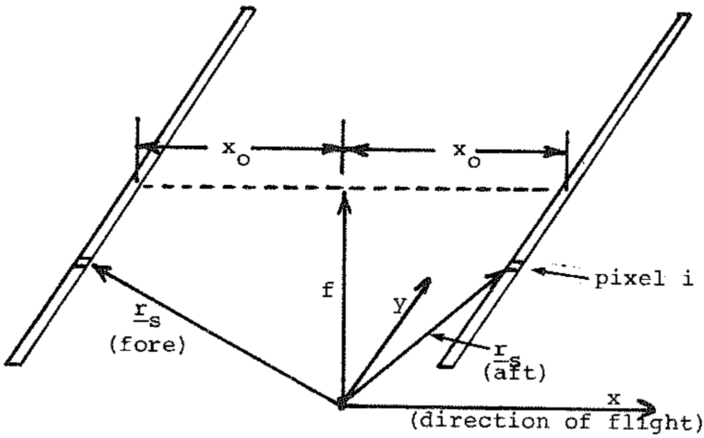


Figure 4: Geometric Correction Geometry

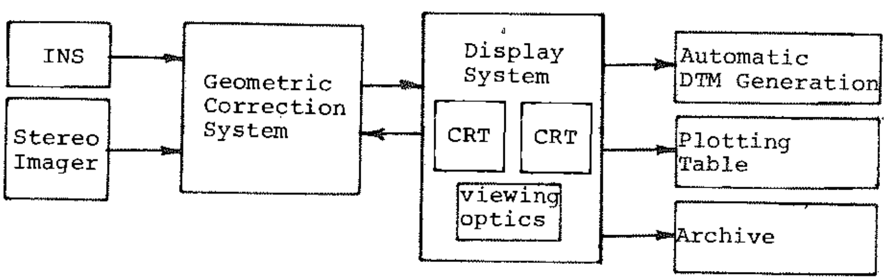


Figure 5: Stereo Line Imager Based Mapping System

UNDERLYING REQUIREMENTS AND TECHNIQUES OF A GEOGRAPHIC DATA BASE

Christine Kinnear and David Meixler
TIGER System Staff
Geography Division
U.S. Bureau of the Census
Washington, D.C. 20233 U.S.A.

ABSTRACT

The Geography Division of the U.S. Bureau of the Census has the responsibility of providing the geographic support for all censuses and surveys conducted by the Bureau. This support includes geographic reference file creation and maintenance, address geocoding, and map production. Traditionally, geographic information to support these functions has not been centralized; instead, each function has been self-supporting. For the 1990 Census and subsequent operations, the Bureau is committed to developing an integrated geographic data base that will support these separate but related functions.

This paper presents the current status of the development of the Topologically Integrated Geographic Encoding and Referencing (TIGER) system. It reviews the system's evolution from theory, through experimental implementation during the 1970's as ARITHMICON, and production testing during the early 1980's as TIGER, to its current state. This presentation includes an overview of the cartographic, address geocoding, and tabulation requirements for the 1990 Census, and a discussion of the goals of the Bureau's prototype geographic data base system. Finally, the underlying data structure of the system is presented, and the techniques used to create and support the current system are described.

KERN MAPS 300

Jonathan C. Hescock
Kern Instruments, Inc.
Brewster, New York, U.S.A.

ABSTRACT

The new interactive graphics station, MAPS 300, is a system module of the Kern Computer-Aided Surveying and Mapping System, CAM.

The Kern CAM System is a rigorous application of the principle of distributed computing. Data acquisition is done with MAPS 100, 200, or 300 stations and stereoplotters or digitizing tables. Data manipulation, merging, editing, and text placing is done with MAPS 300. Data presentation is digital and graphical, with industry-standard drum and X-Y pen plotter, which are supported by the Kern CAM System.

The modular concept allows operation of individual system modules as stand-along work stations, as well as step-wise expansion towards the full Kern CAM System. Moreover, hardware and software service is enhanced through the clear modular structure.

The MAPS 300 Station makes use of the powerful DEC 11/23 computer and the Imlac Series II high-resolution (2048 x 2048 pixels) vector refresh graphics CRT.

MAPS 300 features a full complement of interactive graphics software with many routines especially designed for the surveyor, photogrammetrist, and civil engineer. Some of these features are:

- Vector refresh for instant display of corrections.
- Higher resolution, 2048 x 2048 pixels.
- Interactively edit and transform points, lines, and arcs.
- Interactively edit feature attributes.
- Edit and annotate contour lines.
- Zoom in for fine editing.
- Connect models.
- Place alphanumeric annotation centered, and along straight, angled, and curved lines.
- Find points, lines, and arcs within a file.
- Lock any feature to the cursor for digitizing.
- Audit trail.
- Create graphics files.
- Create symbols and types for repeated placement.
- Create attribute files.
- Merge files.
- Make lines and features parallel.
- Apply scaling and rotation.
- Connect points with straight lines, arcs, and cubic spline interpolation.
- Close curvilinear polygons.
- Easy transportation of data to industry-standard data base management systems.

GEOGRAPHIC DATA PROCESSING IN FORESTRY - DOES IT PAY?

J.A. Benson
Department of Parks and Renewable Resources
Forestry Division
P.O. Box 3003
Prince Albert, Saskatchewan
S6V 6G1

ABSTRACT

Since 1979 Saskatchewan forestry personnel have been examining the practical feasibility of introducing geographic data processing into the manipulation of forestry data, most particulary forest inventory data. From these studies, Saskatchewan has developed a "go slowly" attitude towards the adoption of any full production system. As an interim measure, Saskatchewan will be developing some capability for processing and displaying the coarse raster base presently available for 55% of the inventory mapped data. However, the future looks bright -- the value added to forest land during reforestation efforts renders the incremental cost of creating a vector-style of geographic data base almost negligible.

INTRODUCTION

The first extensive forest inventory in Saskatchewan was initiated in 1947 and completed in 1956; the second was initiated in 1957 and concluded in 1976. In both cases the data base consisted of a set of maps and supporting statistical data summarized for each map. Without onerous re-compilation, the effective geographic resolution of the data base was one map unit. Display of the data was limited to copies of the maps; extensive manual drafting was required to produce thematic displays of the statistical data and seldom was done. At no time during these two inventories was any attempt made to incorporate other forestry data, much less data from other disciplines.

The third forest inventory began in a production sense in 1977 and incorporates considerably more detail than the two predecessors. The effective resolution of the new data base is 0.39 ha thus allowing a myrid of possibilities for data processing and display. It is apparent that by altering the method of data capture and utilizing available geographic data processing technology these possibilities could be enhanced.

This report describes the current practices in Saskatchewan resulting in a computer resident inventory geographic data base, relates past experiences in the investigation of vector-based systems, and looks at the practability of introducing such systems into Saskatchewan Forestry.

CURRENT PRACTICES

The preparation of a forest inventory map follows several sequential steps; in Saskatchewan these are: acquisition of aerial photography, photo interpretation, base mapping, transfer from interpreted photograph to the base map, and error detection and correction. The finished product is a geographically correct map with neat lines following the 6° Universal Transverse Mercator 10,000 m grid, at a scale ratio of

1:12,500. Depicted are combinations of species association (18 classes) stand height (5 classes), crown closure (4 classes), soil drainage (13 classes), soil texture (8 classes), and stand year of origin (18 classes) as well as several non-forested and non-productive classifications; each combination creates a polygon to which a unique number is assigned for that map.

The collection and compilation of field data are beyond the scope of this presentation. Suffice to say that these data relate to the species associations, stand height, and crown closure combinations, are expressed on a per hectare basis, and are applied to each of the mapped polygons.

Thus it is necessary to determine the area of each of the polgons and most efficient to use a computer to effect the linkage to the field data. The method for area determination developed and used in Saskatchewan is as follows:

1. superimpose a dot grid matrix of 16 X 16 per square kilometre on each map;
2. place the map and grid on a digitizing table and register the map to the table.
3. using a pen-type cursor identify the location of each grid-cell by touching the position of each dot, and then the number of the appropriate polygon via a menu; note that dots may be identified singly, or in strings or blocks;
4. verify that the correct polgon number has been entered.

The data set for each map is edited on a mainframe computer and errors corrected on the digitizing table. Two major files are created and linked by map name and polygon number: the master co-ordinate file identifying the location of each grid-cell, and the master stand file containing the attribute list associated with each polygon. Presently data for about 80,000 km are stored in this fashion.

To date, the most extensive use of the data base has been to generate tabular reports for single maps or groups of maps. However, two small sub-systems have been developed that make use of the co-ordinate file. The first incorporates changes to the mapped information onto both the co-ordinate and stand files; the second permits the selection via the digitizing table of portions of maps which then are processed through the reporting systems.

At present, there are no automated systems being used to display information from this data base.

INVESTIGATION OF VECTOR-BASED SYSTEMS

In 1979 Saskatchewan engaged a consulting firm to study the existing forest inventory mapping and data processing systems and to make recommendations for future developments. Following extensive consultation with forest inventory personnel, as well as with personnel from the agencies that use the inventory data, several future requirements were identified by the consulting firm.

<u>Photogrammetry sub-system</u> - If possible, capture the geometric and interpreted forest stand boundary and attribute data directly from interpreted aerial photographs. If practical, this system would result in significant cost reductions by eliminating the manual transfer of interpreted data onto base maps.

Map digitzing sub-system - If the photogrammetry sub-system is not practical, the data must be captured from the finished inventory map. (Probably this sub-system would have been required in any case, particularly to load data for other themes presented in map form.)

Map data handling sub-system - Regardless of the method of data capture, a variety of data handling capabilities would be required including: area measurement, updating, derivation of new maps, corridor analysis, scale change, overlay, interactive browse, and line plotting and attribute labelling.

The consulting firm conducted benchmark tests on five commercially available vector-based systems judged most likely to be able to meet the identified requirements. The results of the analysis of these tests, as well as a description of the entire project, are embodied in "A Study of Forest Inventory Data Handling Systems for the Province of Saskatchewan", prepared by Tomlinson Associates. The single most serious shortcoming of all the systems tested was the inability to efficiently process the quantity of geographic data associated with the test; this shortcoming apparently was related to the format of the stored data and is all the more serious when it is considered that the test data represented only one-sixth of one forest inventory map. Based on the results of the study a vector-based system was not purchased.

Following the study attempts were made to monitor the continuing developments of vector-based systems, primarily through conversations with the various vendors of systems. It became apparent that developments were not progressing as rapidly as had been anticipated. At this time the Lands Directorate, Environment Canada, offered to enhance the Canada Land Data System to meet the Saskatchewan requirements. During subsequent tests the requirement for a photogrammetry sub-system was relaxed, as was the requirement for fully-automated attribute labelling on plotted output. The final production test encompassed an area of about 1600 km^2 and incorporated 114 maps of various themes and scale ratios. Figure 1 illustrates the layers and themes entered and stored on the data base. Through interactive tabular and graphic retrievals, and retrievals plotted on mylar and paper, the Canada Land Data System demonstrated the capability to meet the Saskatchewan requirements. However, it was apparent that due to the sheer bulk of data it would be necessary to partition the Commercial Forest Zone into as many as twenty data sets for efficient retrieval and plotting times.

Since the Canada Land System could meet the identified requirements, an attempt was made to analyze the costs and benefits of installing the system*. An immediate difficulty in the analysis was that few of these future requirements are being met presently and thus cannot be considered essential particularly in terms of graphic displays. Nonetheless, comparisons were made among:

1. maintenance of the existing system, plus the manual preparation of maps of selected generalized themes at a scale ratio of 1:50,000;
2. maintenance of the existing system, plus the automation of grid-cell processing to produce maps similar to the first option;
3. maintenance of the existing system at 50% through-put, plus the utilization of the Canada Land Data System for the other 50%.

---

* Anthony C. Baumgartner

The analysis included a projection over a five-year lifespan and indicated that option 2 was the most cost-effective.

The report describing the Canada Land Data System production test and results* included these recommendations:

1. implement option 2 as the most economical method of preparing generalized forestry maps;
2. use the Canada Land Data System only for those areas on which silvicultural treatments are being carried out; the incremental cost of $0.60/ha for using the system for detailed land base record keeping is small in comparison to the costs of silvicultural treatments which range from $80/ha to $400/ha. The equipment and associated infrastructure required for option 2 would be sufficient for implementing this recommendation.

CONCLUSION

Forest and resource management decision making is becoming more complex and the timing of those decisions more urgent with the increasing pressure for use of the land base. Those "non-essential" future requirements for geographic data processing are becoming essential and there is no doubt that future cost/benfit studies will come down in favour of more sophisticated systems. The decreasing real price of computing power, data storage facilities, and plotting devices will hasten that process. Forestry in Saskatchewan cannot help but benefit from improved geographic data processing.

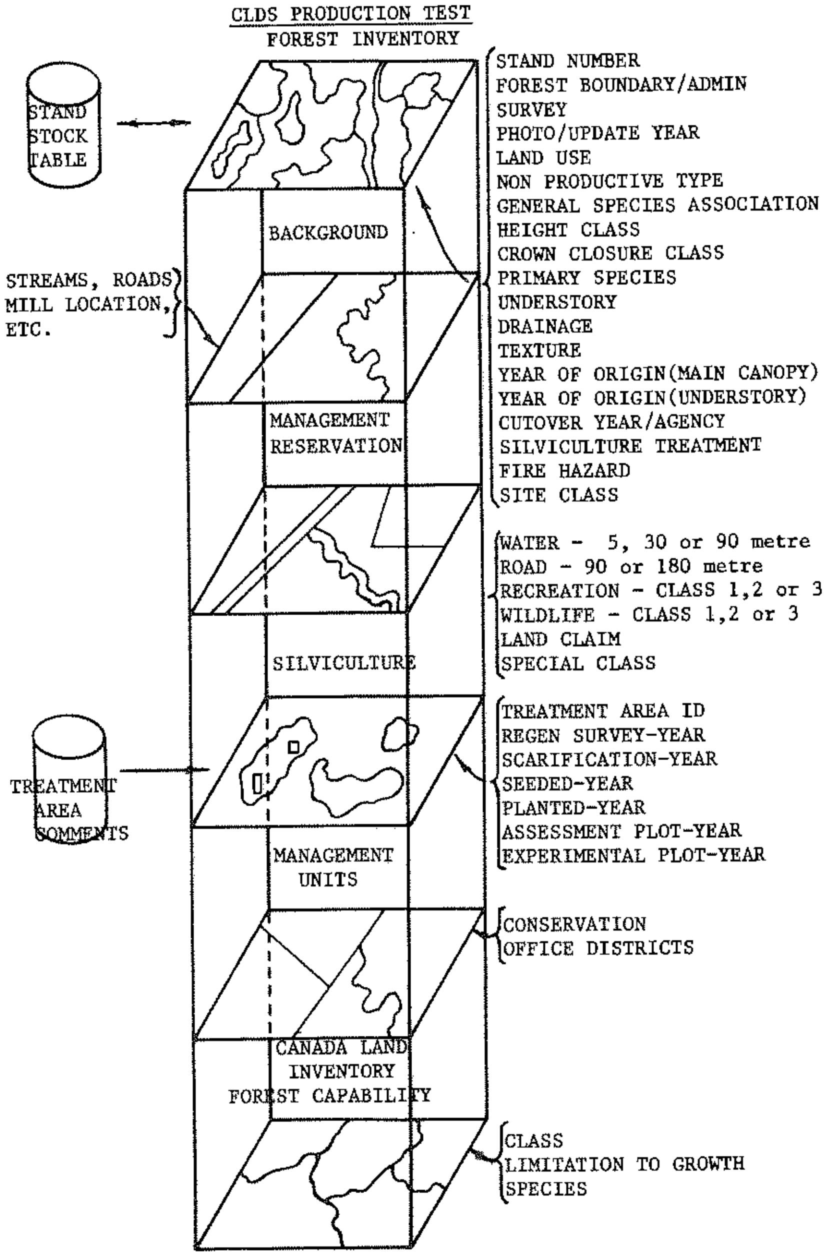
REFERENCES

Baumgartner, Anthony, C. 1982, Economic Analysis of Geographic Data Processing Alternatives: internal report, Forestry Division, Saskatchewan Department of Parks and Renewable Resources.

Fisher, T.A. 1982, A Study of the Geographic Data and Data Processing Requirements of the Forestry Division: internal report, Forestry Division, Saskatchewan Department of Parks and Renewable Resources.

Tomlinson Associates 1980, A Study of Forest Inventory Data Handling Systems for the Province of Saskatchewan: Forestry Division, Saskatchewan Department of Parks and Renewable Resources.

* T.A. Fisher


CLDS PRODUCTION TEST
FOREST INVENTORY
STAND STOCK TABLE
BACKGROUND
STREAMS, ROADS, MILL LOCATION, ETC.
MANAGEMENT RESERVATION
SILVICULTURE
TREATMENT AREA COMMENTS
MANAGEMENT UNITS
CANADA LAND INVENTORY FOREST CAPABILITY
STAND NUMBER
FOREST BOUNDARY/ADMIN
SURVEY
PHOTO/UPDATE YEAR
LAND USE
NON PRODUCTIVE TYPE
GENERAL SPECIES ASSOCIATION
HEIGHT CLASS
CROWN CLOSURE CLASS
PRIMARY SPECIES
UNDERSTORY
DRAINAGE
TEXTURE
YEAR OF ORIGIN(MAIN CANOPY)
YEAR OF ORIGIN(UNDERSTORY)
CUTOVER YEAR/AGENCY
SILVICULTURE TREATMENT
FIRE HAZARD
SITE CLASS
WATER - 5, 30 or 90 metre
ROAD - 90 or 180 metre
RECREATION - CLASS 1,2 or 3
WILDLIFE - CLASS 1,2 or 3
LAND CLAIM
SPECIAL CLASS
TREATMENT AREA ID
REGEN SURVEY-YEAR
SCARIFICATION-YEAR
SEEDED-YEAR
PLANTED-YEAR
ASSESSMENT PLOT-YEAR
EXPERIMENTAL PLOT-YEAR
CONSERVATION
OFFICE DISTRICTS
CLASS
LIMITATION TO GROWTH
SPECIES

LONG-TERM PLANS FOR THE GEOGRAPHIC AND CARTOGRAPHIC SUPPORT TO THE U.S. BUREAU OF THE CENSUS

Frederick Broome, Timothy Trainor, and Stephen Vogel
Mapping Operations Branch
Geography Division
U.S. Bureau of the Census
Washington, D.C. 20233 U.S.A.

ABSTRACT

The United States Bureau of the Census has embarked upon an ambitious effort to automate its geographic support for the 1990 Decennial Census. All relevant geographic information about an area will be recorded in a single computer file. This system, when complete, will permit the Bureau to perform the three major geographic support functions--assignment of geographical codes to residential and business addresses for data collection, provision of a geographic structure for statistical tabulation and publication, and automatic production of cartographic products to support data collection and publication work. Geographic support previously has been provided through a mix of automated and manual means involving multiple, unconnected computer files, hundreds of clerks, and tens of thousands of maps. The results were many problems. Because all products will be produced from one file, the possibility of errors similar to those which occurred previously will be vastly reduced, if not eliminated.

This paper describes the long-term plans of the U.S. Census Bureau's Geography Division for the 1990 Decennial Census geographic support system. The paper reviews the preliminary design for the computer file, its structure, and the mathematical theories underlying it. The paper reports on the development of the functional requirements for the file and the system and the establishment of content standards for both spatial and non-spatial data. The file is called the Topologically Integrated Geographic Encoding and Referencing file, or "TIGER."

DATA STRUCTURES

Geodesic Modelling of Planetary Relief 186
Geoffrey Dutton

A Structured Expert System for Cartography Based on the Hypergraph-Based Data Structure (HBDS) 202
François Bouillé

Building a Hypergraph-Based Data Structure: The Examples of Census Geography and the Road System 211
Robert Rugg

Multiple Data Structures in a Regional Data Base 221
Paul Wilson

Adaptive Grids for Geometric Operations 230
Randolph Franklin

Terrain Approximation by Triangular Facets (Abstract) 240
Albin Tarvydas

Reduction of Digital Aerogeophysical Data to a Linear Model (Abstract) 241
Michael Holroyd

GEODESIC MODELLING OF PLANETARY RELIEF

Geoffrey Dutton
Laboratory for Computer Graphics
and Spatial Analysis
Graduate School of Design
Harvard University
Cambridge, MA 02138 U.S.A.

ABSTRACT

A method for assembling and managing global terrain data is presented, the Geodesic Elevation Model. Derived from concepts in geometry, geography, geodesy, applied mathematics and computer science, GEM is designed to digitally archive and access measurements of points given in latitude, longitude and elevation from any source, by embedding them in a regular, polyhedral data structure. To do this, the model recursively tessellates a regular solid, either an octahedron or an icosahedron, into equilateral triangular facets. Spot measurements are encoded by successive approximation, mapping a given geodetic location to proximal centroids of nested triangles. As encoding proceeds, a new vertex appears at the center of each existing facet; an elevation code for it is entered in a linear tree, an estimated coordinate which locally wrinkles the polyhedron. The more times this takes place, the better is the approximation: each such step of encoding triples the number of facets, and diminishes horizontal and vertical error by the square root of three. As the structure is a regular geodesic grid, its horizontal coordinates are implicit by their ordering. Elevations alone are stored, using 1-bit flags quantizing height changes, triangulating the enclosed terrain with less than one bit of data per facet. Reconstruction of the data yields estimates of longitude, latitude and elevation anywhere on the planet, along with the error of estimate. Consequently, the slope, size and aspect of facets can be derived at any level of precision required, up to the limit of detail encoded for their neighborhoods. Beyond this, if desired, fictitious detail can be fractally synthesized, landforms resembling features above them in the hierarchy, smoothing the surface simultaneously. Local regions (small initial facets) can be encoded and stored independently, then subsequently merged at will to assemble larger terrain models. As measurements accumulate in a GEM database, superfluous and erroneous data are rejected with increasing frequency, due to self-calibrating nature of the ensemble. Were sources of data and motivation sufficient, the relief of the entire Earth could be uniformly encoded in GEM format at a horizontal resolution of less than one kilometer, with a vertical precision of several meters, on a single disk volume.

Digital Terrain Models (DTMs) are becoming increasingly important in civil engineering, regional analysis, military operations and topographic mapping. Over the past fifteen years or so, a number of data structures and formats have been developed to encode topographic surfaces (Mark, 1977), some of which are mathematical (usually coefficients of polynomial or trigonometric series, globally approximating relief as smooth surfaces or locally as patches).

Perhaps the purest example of a mathematical terrain model, and in many ways closest in spirit to the model presented in this report, was formulated between 1908 and 1922 by Prey (1968). In that model, the shape of the earth was represented by spherical harmonic equations to order sixteen. To have refined his model further would have exceeded the computational power and demanded better data than were available in Prey's time.

Since then, not only has the development of computing technology vastly increased our command of techniques, this revolution has unloosed an avalanche of spatial data, burying our abilities to catalog, verify, analyze and apply it. The work reported here is, like Prey's effort, an attempt to bring some unity to the chaos of DTM technology and data.

Gridded Terrain Models

The most common way to represent geographic data is to accumulate observed or interpolated point data into fixed rectangular grids or irregular triangular meshes. The former type are by far the dominant form for DTM data, as they can be compiled nearly automatically using analytic stereoplotters to scan stereo pairs of aerial photos, correlating features to compute elevation profiles as a dense raster of points. DTMs produced by such machines (after resampling and editing) are publicly distributed by the U.S. Geological Survey; these are available for a growing number of 7.5-minute quadrangles, representing their terrain with about 170,000 grid points spaced at 30-meter intervals. Quality control of such data is problemmatic, as the accuracy of the output from the analytic stereoplotters varies with terrain type and land cover, and in particular is affected by the presence of water bodies and man-made structures.

Triangulated Irregular Networks

The principle alternative to rectangular arrays for storing DTM data is the Triangular Irregular Network (Peucker et al, 1977; Males, 1978; Gold, 1978), or TIN. In this model, elevations are digitized from maps manually, selected to represent critical features of the surfaces to be encoded, such as peaks, pits, passes and breaks in slope. These points are then triangulated, either manually or analytically (Fowler and Little, 1979), yielding a varying network of triangular facets fitting the terrain, containing most of the information about the surface with relatively little input data. While some thought must be given to selecting and connecting the initial spot elevations, the resulting model is more compact and useful than gridded DTMs, principally because of the properties of triangles and the networks they form. The main drawback to TINs is the complexity of the data structures and programming strategies needed to manage and apply them. Furthermore, while all implementations of TINs are conceptually equivalent, their structures differ sufficiently to make data transfer from one system to another difficult. Transferability issues also arise for gridded DTMs, but usually due to differing data formats (ordering and character encoding conventions) rather than because of any fundamental differences in data structure.

Accuracy Issues

Despite their different constructions, there are some common aspects to and shortcomings of the two models just outlined. Their principal similarity is that both grids and TINs are designed to encode planar coordinates for relatively small areas. That is, the horizontal coordinates (which are explicit in TINs but implicit in grids) are almost always cartesian, and scaled to whatever map projection was used to compile the source maps or photographs from which they are digitized. In too many cases, the projections are not documented, leading to difficulty when adjacent terrain models are merged or when particular ones are modified.

Another set of limitations relate to the accuracy of DTM data. Gridded data has an implicit limit to horizontal precision (its Nyquist frequency), which is nominally uniform throughout the grid. In practice, however, gridded surfaces may be produced by interpolation procedures from scattered observations. The resultant precision is thus variable, but its magnitudes are hidden, unless one has access to the source data.

Vertical precision, likewise, may or may not be uniform throughout a grid, depending on the methods and sources used to compile it. Moreover, a structural interdependence exists between it and horizontal precision (sampling density), demonstrated in Dutton (1983). If grid cells are large they may contain considerable amounts of height variation, so that the value assigned to each is only an estimate (high, low or average) of heights within the cell. To ascribe high precision to the elevations assigned to grossly-sampled cells is thus rather pointless unless the terrain is generally smooth. Therefore, the amount of memory needed to represent the height of a grid cell grows larger as sampling density increases. Specifically, the number of bits needed to encode each cell is proportional to the logarithm of the number of cells.

Precision in TINs is subject to similar constraints, as each control point may vary in how well it represents conditions in its neighborhood. In general, precision will vary inversely with the spacing of observations, but need not (and probably will not) be the same at each control point, due to source errors, variations in operators' performances and triangulation decisions. While each coordinate and face in a TIN can be labelled to document its presumed precision, this is never done in practice. To do so would erase much of the storage efficiency enjoyed by TINs.

In editing gridded surfaces and TINs, gross errors can be detected automatically by algorithms which identify drastic changes in slope or linear artifacts. Visual inspection is still the best way to achieve quality control, but only errors that result in discernable patterns are likely to be rectified. The overall fidelity of the data is difficult to assess without detailed information of how the source data were collected, edited, reproduced and (sometimes) interpolated. The datasets themselves are completely indifferent to the quality of the information they contain, and this implies that they will usually contain errors which will persist without notice, but unfortunately not without consequence, indefinitely into the future.

FORM

In the following sections, a structure for encoding terrain is presented which differs significantly from both grids and TINs in both the vertical and horizontal components. Like a grid, elevations alone are explicitly encoded, in a regular mesh; like a TIN, all cells are triangles with identifiable faces, vertices and edges. Unlike either, however, the model is designed to be planetary in scope, capable of accepting observations from any location on earth (or whatever planet it represents), storing them in geodetic (spherical) coordinates. Furthermore, if observations are properly labelled as to their horizontal and vertical precision, the model will encode each to the appropriate tolerance and no further. When they are sufficiently in error, input data can be rejected by the model automatically. These and other useful properties are achieved with surprising economy; each facet encoded requires storage of <u>less than one bit</u> of data, making the model at least an order of magnitude more compact than either grids or TINs. This method of encoding planetary relief has been named Geodesic Elevation Modelling (GEM). Some of its properties are summarized in Table 1, in comparison with gridded and triangulated data structures.

Table 1

COMPARISON OF GRID, TIN AND GEM STRUCTURE AND PROPERTIES

DTM:	GRID	TIN	GEM
BASIS	Raster	Landforms	Planet
FORM	Cartesian	Triangulated	Polyhedral
TOPOLOGY	Implicit 2-d	Explicit 2-d	Implicit 3-d
SAMPLING	Uniform	Irregular	Hierarchical
PRECISION	Fixed	Variable	Convergent
CONTENTS	Elevations	x,y,z,pointers	Diff. Codes
STRUCTURE	Array	linked lists	dual trees
STORAGE/CELL STORAGE/MAP	medium high	high medium	low low
ACCURACY	standardized	uncertain	verifiable
COMPLEXITY	low	high	medium
COVERAGE	local	regional	global
EDITING	tedious	complex	automatic

Geodesic Tessellation

A geodesic structure can be generated from a polyhedron (usually an icosahedron* or portion of one) by regularly subdividing it in several well-defined ways. In two of these methods, the so-called "triacon" and "alternate" breakdowns (Popko, 1968), a higher-frequency grid of triangles is created by connecting either the centroids and vertices of triangles (triacon, fig. 1b) or their edge midpoints (alternate, fig. 1a). The two procedures yield 60 and 80 triangular facets, respectively, from the original 20 faces of the icosahedron. Each of these new triangles can subsequently broken down in the same manner, each time tripling or quadrupling the number of faces in the structure. Eventually, the triangles grow quite small and the figure begins to closely approximate a sphere. Ten orders of subdivision or less is about the limit for engineered structures; edge members grow quite short, yet of slightly (but critically) differing length, posing tolerance problems in manufacturing them and their connectors. Computational models, however, can be made of such structures without encountering such problems, at least to one part in several million.

Polyhedral Projection

While geodesic structures are normally regarded as minimal space-enclosing ones, they can also be thought of as models of surfaces which are, topologically speaking, spheres. Fuller and Sadao (1982) have published a "globe" which represents the earth as an icosahedron, the so-called "Dymaxion" projection. Although ingenious (it can be cut and folded from a single sheet of paper) and informative (it offers many untapped possiblilities for thematic mapping), the Dymaxion projection is not widely used, being regarded by mapping professionals as a cartographic curiosity. While the Dymaxion Map distributes its projection errors predictably, its distortions are concentrated at vertices and arbitrarily located, depending on the orientation of the icosahedron with respect to the earth's axes. Where accuracy must be predictable, then, the Dymaxion projection is not the one of choice.

These limitations recede, however, as one proceeds to subdivide the icosahedron into smaller triangles. With each successive division edge lengths are reduced by half (for the alternate breakdown) or by the square root of three (triacon breakdown) and the polyhedron grows more spherical. Its facets lie closer to sea level and projection distortions relax. After only five (alternate) breakdowns, the original twenty icosa faces blossom into 20,480, with vertices arrayed 2.25 degrees apart. This rate of expansion can be continued to make the polyhedron indistinguishable from a sphere for any cartographic purpose.

Fuller uses the icosahedron as the basis for most of his geodesic structures, a choice which is both theoretically and pragmatically justified for engineering purposes. But despite the appeal of icosahedral forms for space-frame structures, they may not in fact

---

*An icosahedron is a polyhedron having 20 equilateral triangular faces joined along 30 edges which connect each of its 12 vertices to 5 neighboring ones. It is the largest regular platonic solid which is convex (not stellated). This figure is the basis for the construction of a large family of space-frame structures developed by R. Buckminster Fuller, his colleagues and followers. For a comprehensive statement of the principles governing these constructions consult Fuller, 1982.

FIGURE 1a

The "Alternate" Geodesic Hierarchy

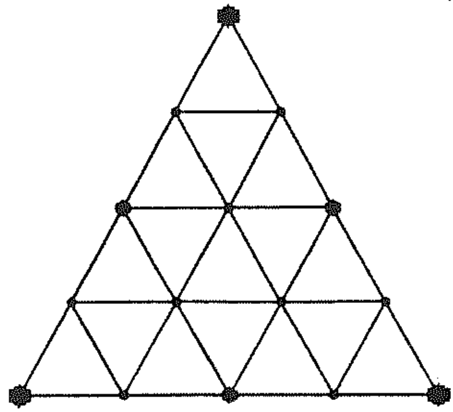
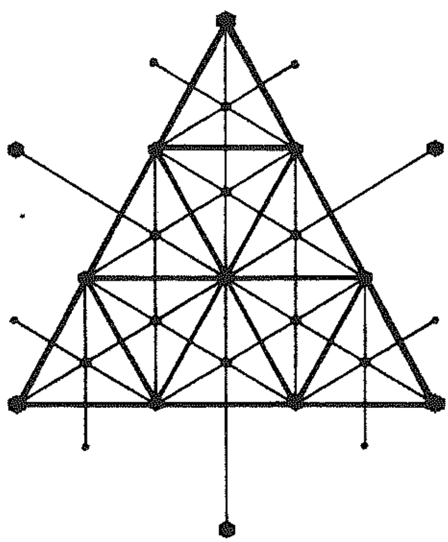


FIGURE 1b

The "Triacon" Geodesic Hierarchy



provide an equally optimum basis for modelling geodetic relief. The facets of an icosahedron are tilted with respect to the equator and prime meridian, complicating computations of their geodetic coordinates.

An octahedron (having 8 triangular faces, 6 vertices and 12 edges), on the other hand, can be aligned to cardinal points, and this property leads us to regard it as a more appropriate polyhedron upon which to structure a geodesic data base. If oriented so that the polar axis passes through two opposite vertices and so that the prime meridian and the equator intersect at another vertex, the octahedral kernal acquires certain useful properties. First, all "baselines" of the initial facets and their subdivisions are parallel to the equator. Furthermore, the difference in latitude or longitude between vertices is divided by three with every pair of subdivisions. This yields an isometric graticule which grows finer with each breakdown, three sets of "standard parallels", two of which are oblique. Each face is semi-uniquely defined by the three parallels which intersect about it (two such faces are formed at antipodes). One potential disadvantage, however, is that the shape of facets will vary from nearly equilateral (near octahedron face centers) to right spherical triangles (at octa vertices), and consequently their areas will differ. The size and shape of any facet can always be computed, and in any case their variation does not make the model any less useful for storage and retrieval of elevation data.

FEATURES

In the following sections, one implementation of a Geodesic Elevation Model is described. This model is based on the Triacon Breakdown of an Octahedron, ignoring other geometries, but not dismissing them as inappropriate. Were other breakdowns of other figures to be employed, the logic of the model would be very similar; certain parameters (such as the rate at which facets multiply) and data structures (memory addressing strategies) would be altered, and certain convenient properties might be foresaken.

Like any digital terrain model, GEM has three major related elements:

1. Horizontal Organization
2. Vertical Encoding
3. Data Structure

Its horizontal organization is that of the octahedral triacon breakdown, a regular tessellation which, like a raster-encoded image, needs no horizontal coordinates: location is implied by position in the data structure. The two alternating hierarchies of the triacon provide a triangular matrix of control points regularly arrayed across the surface of the planet. Consequently, only vertical information need be contained in a GEM file, encoded as bit-flags which signal the elevation change at each control point. The flags describe the direction of elevation change, but not its magnitude. The amount of vertical movement is, analogously to horizontal offsets, given by the position (depth) of facets in the hierarchy. This relatively unexplored method of elevation modelling has been named DEPTH, for "Difference-Encoded Polynomial Terrain Hierarchy" (Dutton, 1983). Its effect is to approximate elevations vertically to a similar extent that locations are approximated by triangular breakdowns. Each level of a DEPTH hierarchy consequently encodes more control points with greater precision than the levels above it. Table 2 enumerates this hierarchical progression to twelve levels, illustrating the asymptotic approximation of a sphere of

Earth radius starting from an octahedron fitted around it.

Table 2

"TRIOCTACON" TESSELLATION OF A SPHERE WITH A RADIUS OF 4000 MILES

LEVEL	NO FACES	NO POINTS	NO EDGES	EDGE LEN	FACE AREA	VERT ERR.
1	8	6	12	6531.969	18475188.	1952.134
2	24	14	36	4046.708	7090952.	634.494
3	72	38	108	2390.866	2475205.	211.091
4	216	110	324	1390.952	837770.	70.350
5	648	326	972	805.109	280679.	23.449
6	1944	974	2916	465.224	93718.	7.817
7	5832	2918	8748	268.673	31257.	2.605
8	17496	8750	26244	155.133	10421.	0.869
9	52488	26246	78732	89.569	3474.	0.290
10	157464	78734	236196	51.713	1158.	0.097
11	472392	236198	708588	29.857	386.	0.032
12	1417176	708590	2125764	17.238	129.	0.011

This series of breakdowns converges rapidly. Its faces multiply by powers of three, reducing the triangles at the 12th level to less than 130 square miles apiece. By this level, vertical error is such that the center of each triangle is about 50 feet from the spherical surface. As implemented in GEM, however, the rate of spherical convergence is reduced from a ratio of three to the square root of three.

GEM's archival and working data structure represents vertices as two parallel sequences of levels in dual nonary trees. In each hierarchy every non-terminal triangle divides into nine, each with one third the edge length of its parent. Because all descendant nodes are represented in the tree, there is no need for pointers; the tree is laid out as a series of arrays, each three times (more or less) longer than the last. Access to values is then through computing an address from the coordinates of a centroid and its depth in the tree. Similar data structures for quadtree hierarchies are in use, sometimes called "linear trees" (Gargantini, 1982). In the absence of list pointers, all operations on such structures proceed sequentially, top-down.

<u>Horizontal Encoding</u>

GEM's triacon grid has the useful property that centroids of triangles mark elevation nodes, which then serve as vertices for triangles in the next lower level. Every terrestrial location can be approximated within <u>epsilon</u> distance units by a specific sequence of triangulations converging about it. The primary facet can be one of the initial polyhedron, or one of its divided facets. In either case, the same process is used to continue breakdown. The procedure generates a series of partially or completely nested triangles, the numbering of which provides both a geocode and a key for memory addressing, as well as the vertex coordinates of each triangle in turn.

To characterize the recursive strategy of approximating locations via successive triangulations, the term <u>trilocation</u> has been coined. Use of this neologism will simplify subsequent discussions, as should its syntactic variations, such as <u>trilocate</u>, <u>trilocated</u> and <u>trilocal</u>. Computations for trilocating can be very simple, due to the regularity of the geodesic structure. Algorithmically, to trilocate point [pq], given that it is bounded by vertices [p1,p2,p3], perform:

FIGURE 2a

GEODESIC SPATIAL SEARCH: The TRILOCATION Procedure

The small cross marks a location to be found: Six Trilocations are shown, converging to within 4% of initial edge length. The "best" estimate at each stage is the center of the current triangle.

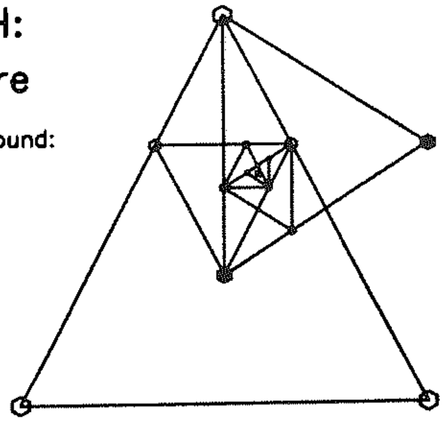
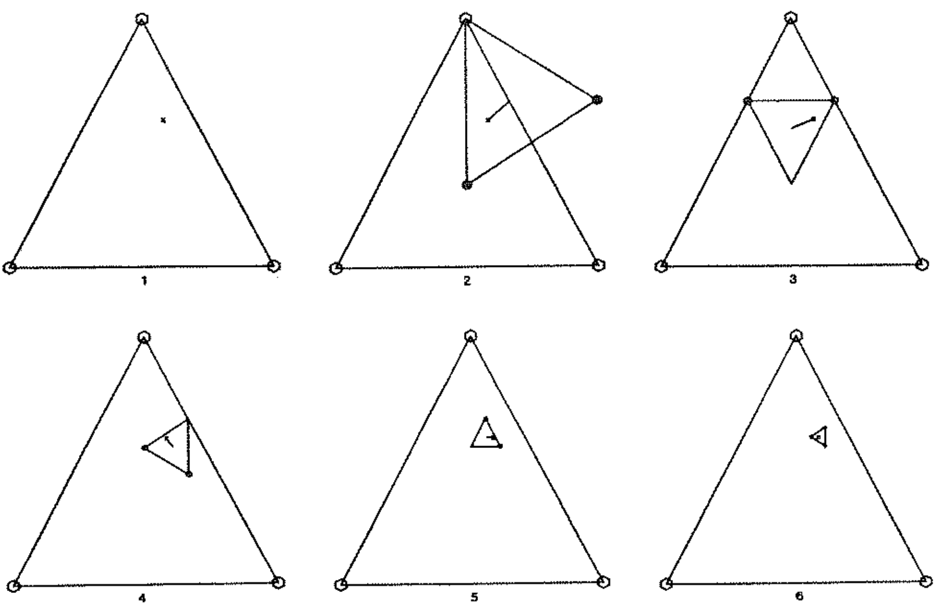


FIGURE 2b



```
PROCEDURE TRILOC (Real: pq,p1,p2,p3,p1new,p2new,p3new,center,epsilon)

;Converges a triangle around a point to a new one one-third as big;
;Vertices are pairs of planar or spherical coordinates.
 BEGIN:
  Determine squared distance from [p1] to [p2]
  If less than [epsilon] squared then return; we're close enough
  Determine squared distances from [pq] to [p1], [p2] and [p3]
  Sort them, ordering points as [pnear], [pmid] and [pfar]
  Compute vertices of convergent triangle containing [pq]:
    [p1new] ::= [p1 + p2 + p3] / 3; centroid of current triad
    [p2new] ::= [pnear]; closest old vertex
    [p3new] ::= [pnear]+[pmid]-[p1new]; centroid of nearest neighbor
    [center] ::= [p1new + p2new + p3new] / 3; best estimate of all
 END.
```

Each time TRILOC is called, the edge length of resultant triangles grows smaller by the square root of three. This creates an alternating sequence of triangles, in which every level is fully contained within the level two steps above it, and has nine times as many triangles, each with one-third the edge length. This pattern of breakdown is shown in figure 1b, with odd-numbered breakdowns drawn with bold lines. The trilocation procedure is diagrammed in figure 2, showing the approximation of a location in six steps. Note that while each trilocation generates bounding vertices, it is the center of that triangle which marks where vertical information is encoded at that level.

Vertical Encoding

Considered three-dimensionally, the two alternating triacon grids can be regarded as constituting a pair of concentric polyhedra. The odd levels, those of the initial octahedron and its subdivisions, generate an object which shrinks slightly with each subdivision. Those of the other form start as the dual of the octahedron, a cube. As this network is subdivided the volume it contains grows larger. If the radius of the octahedron is initially set to be somewhat larger than the earth (actually, 7.75 percent larger), the radius of the cube (where the radius is the distance from its center to any vertex) will be smaller than Earth's by the same percent. In the process of subdividing, the vertices of both figures tend to converge to Earth radius, eventually approximating spheres which are almost coincidental.*

What DEPTH encoding does is to add information to each vertex in each level of both hierarchies which causes this process of convergence to depart from sphericity. Flagging on a DEPTH code for a vertex can cause it to locally "dimple" or "pimple". Because two networks are involved, only one kind of wrinkling occurs at each. That is, the

---

*The duality of the cube and octahedron is central to Fuller's geometrics. One of his most interesting discoveries is a construction known as the "jitterbug", which demonstrates the instability of the cube-oct form, which Fuller terms "vector equilibrium". The jitterbug swings ambivalently from from an octahedral state through a cube-octahedron stage before collapsing again into a re-aligned octahedron, pumping in and out in alternating cycles. GEM's geometry comes quite close to being a high-frequency jitterbug.

initial (octahedral) network, being larger in radius than the Earth, will contain the pimples (peaks and ridges), while the cube-based network, smaller than the Earth initially, will encode the dimples (pits and courses). Each vertex is assigned a single bit of storage, and will wrinkle if that bit is set, but will maintain a smooth curvature if its bit is not set. The direction of the wrinkling is fixed for each network (which is why only one bit is needed to encode it). The magnitude of the wrinkling is constant for all vertices at a given level, and diminishes geometrically (by root three) as triangles grow smaller, the dual networks converging.

DEPTH represents (discrete) changes in elevation rather than (continuous) heights. It works by comparing the value of an elevation to be encoded with the prior estimate for it (which can be arbitrary initially); if the difference between the actual and the estimated heights is greater than the tolerance in effect for that level, a flag is set to indicate that the estimate is to be raised or lowered by the current step size (which will be related to the tolerance). Moving down the hierarchy, step size gets smaller, halting the encoding when the displacement becomes less than some stated vertical error tolerance; this reflects either how precise the source data is presumed to be or how accurately one wishes to encode it.

Formally, this constitutes a polynomial series (Dutton, 1983, eq. 5) for each vertex, generating a weight for each term as a dummy variable. The weights are then multiplied by the step size for their level. Terms with weights of unity are added (if even) or subtracted (if odd); if a weight is zero, its term is ignored and displacement is toward sphericity. GEM's data structure is simply a way of ordering these two series of weights, bit after bit.

Elevation flags can be grouped into files containing fields for a specified number of levels. It is probably advantageous (and certainly simpler) to make the hierarchical depth (hence size and layout) of each file identical. Within a file, vertices can be clustered by level and arrayed contiguously, ordered by vertex number. This number, in essence a mathematical geocode, is assigned by counting vertices in a uniform way at each level. When finding the central point of a facet via trilocation, its location code can also be computed, then converted into a (byte) memory address for the linear tree. To be most useful, the vertex numbering scheme should generate unique geocodes directly from latitude, longitude and level. The numbering problem has not been fully solved for GEM, and suggestions in this regard are welcomed.

Quality Control

At one time every introductory lecture on data processing seemed obliged to note the acronym GIGO: Garbage In, Garbage Out. This rubric is usually a preface to the assertion that computers are mechanical, obedient and stupid, their software incapable of evaluating the integrity of information being manipulated. For most applications programs and database systems, this still is more or less the case. While each variable and parameter may have a defined set or range of valid values, incorrect ones can still penetrate such coarse filters, biasing tabulations and statistics, polluting databases.

As sets of measurements go, spatial data are distinctly corruptible. Map data has at least three dimensions (spatial location and value), each of which is subject to several sorts of error at various stages of compilation. The passage of time tends to blur the significance of most measurements, especially those that are catalogued without citation of sources and estimates of accuracy. Moreover, it is often more trouble than it is worth to attempt to overlay or merge map files compiled at different times, scales and in different formats, seriously hampering their utility for analysis and planning applications.

Validation. Despite having a simple structure and lacking redundancy, a GEM database has considerable capacity to detect and reject suspicious data, provided that it is initialized with accurate measurements which are fairly widely distributed, rather than being clustered in a few locations. Furthermore, all coordinates input must be explicitly tagged with estimates of horizontal and vertical accuracy to be acceptable. These error terms then determine the lowest level in the structure to which encoding can proceed. At some stage, vertical error will exceed the magnitude of vertical displacement, making it no longer possible to confidently assign a DEPTH code. Similarly, once horizontal error exceeds about one-half the distance between neighboring vertices, it is no longer possible to locate an observation in a particular facet. This means that less certain measurements are limited to the higher reaches of the hierarchy, while more precise ones dive deeper into it.

Tuning. Once installed in the tree, a spot measurement fixes the heights of all the vertices it has visited during its trilocation. Should another measurement subsequently be inserted which follows the same path (for any number of levels), because it is located near the first one, it must generate DEPTH codes which agree with those already in place. If, during this process, a DEPTH code is generated which conflicts with one already stored for a facet, the accuracy of one or both measurements is called into question. A general rule for resolving such contentions can be that the measurement having the smaller error terms (hence a greater depth of encoding) wins. Normally, one tries to insert the most reliable data into a GEM database first, so that they can referee subsequent measurements in their neighborhoods. For example, one could pin down the structure by first encoding known survey monuments, mountain peaks and coastlines (where tidal inundation is not extreme) before entering less verifiable elevations. If, however, a subsequent measurement has greater precision than one already encoded, it should logically supersede it. This can be done, but may require a number of values inferior to the conflicting vertex to be modified in the process.

Adjustment. To verify whether a candidate elevation is more precise than any measurement already stored within a facet where conflict occurs, one searches for set elevation flags inferior to the facet. If none are found, the value can be entered without altering any other data. Otherwise, flags will have to be re-set to maintain the heights currently encoded for them. A refinement is to adjust elevation codes outward from the node just encoded, interpolating values for neighbors at levels between their current depth and that of the new measurement.

Interpolation. The final triangulation estimates elevations at the places where the last two levels cross (at the middle of their edges). It is only at these locations that the two hierarchies meet. All other vertices are either in one network or another. The height computed will be the average of the four vertices which define the intersection. The result generates a network of right triangles defining the surface of the planet. Two such triangles are generated for each one in the lowest level encoded, doubling spatial resolution.

Like any array storage structure, GEM provides a place for everything, with everything in its (approximate) place. As GEM storage is hierarchical, the size of the place a thing occupies can vary with its depth. Space must be pre-allocated, both in memory and in files, for all vertices throughout a breakdown, for as many levels as one intends to encode. While only one bit per vertex is needed, the number of them grows geometrically. Specifically, in a hierarchy of N levels of triangles, there will be 3**(N-1) ultimate vertices (leaves in the tree), and (3**N)/2 vertices overall, requiring (3**N)/16 bytes to fully represent the structure. To give examples, six levels will require just 46 bytes of data space, but 12 levels demand 33,216 bytes, and to hold 18 levels 24,213,781 bytes must be committed.

Reserving storage is cheap to begin with, but clearly grows prohibitively expensive beyond 12 levels or so. One thus must accept a tradeoff between a unified, inflated structure and a limited, manageable one. At the expense of slightly greater algorithmic complexity and processing time, the geodesic hierarchy can be segmented into two or more orders, avoiding the inefficiency of storing unencoded areas. This means that each initial facet is rooted in a first-order file which contains L levels of detail. Another L levels of further detail are then available from a group of second-order files: as many of these may exist as there are leaves in the first order tree, although none need be created unless elevations are encoded for the facets they contain. When encoding or searching beyond the first order, the appropriate second-order file is created or read into memory, doubling the amount of data held there.

The type, size and format chosen for GEM files depend on many factors, few of which are intrinsic to the model itself. Fast direct access is desirable, but properly buffered, sequential files can be highly efficient. For simplicity of access and update, fixed block random files may work best. As already mentioned, file size can inflate enormously if one attempts to encode many levels of the tree at once. In order to tune GEM to virtual memory environments, data should be blocked into page-sized records, such as 256, 512 or 1024 bytes. Given a 512-byte page size, it should be possible to store eight levels of triangles (more than 6500 faces) within it, accessing them in a single read instruction. This tessellates the Earth into facets roughly 3,500 square miles in area, defined by vertices spaced about 100 miles apart. Eight such pages, 4096 bytes in all, would thus represent the earth's shape as a first approximation.*

To then extend the hierarchy another eight levels, each of the 52,488 facets of the first order (8 sets of 6561 facets) would generate 6561 facets of its own, requiring storage of one page of data apiece. This amounts to nearly 350 million facets for the whole planet, and uses slightly less than 27 million bytes to store their DEPTH codes. Each 16th-order facet intersects with a 15th-order one in the final

----

*These first eight levels of detail do not contain much information about the relief of the Earth, but do reflect its shape. If not encoded, this hierarchy assumes spherical proportions. If encoded with geodetic data, however, the 52,488 facets can model the geoidal shape quite adequately. Construction of such a model geoid alone might make the investment in developing GEM worthwhile.

interpolation, yielding about 690 million ultimate triangles. Typically, they cover about 0.075 square miles in area, and have vertices spaced about 0.3 miles apart.

Programmability. Despite that (or because) organizing and manipulating GEM data involve such primitive operations, not every programming language makes them easy. As data are arrayed by bits and addressed by bytes, using languages such as FORTRAN, BASIC and even PASCAL invites hacking byzantine, inefficient code. More polymorphous ones, including C, FORTH, LISP and of course Assembly Code, support GEM operations much more efficiently. In the current research effort a hybrid approach is being attempted, using C to code primitive operations and FORTRAN for higher-level procedures. In our environment (Vax 11/780, VMS), this is accomplished easily, without recourse to Macro coding.

FORWARD

Viewed from space, the Earth assumes the form of a beautiful, varigated bubble. In our efforts to measure and model its topography, we forget that as rugged as landforms may appear to us here, they are merely its texture. Were we able to compress it to be one meter across, the Earth would feel as smooth as the proverbial Billiard Ball: Mount Everest would protrude just half a millimeter above the spheroid. GEM offers a way to model the form of our planet or others, gracefully and economically, and as completely as one is prepared to measure and encode its surface.

In the course of several years, GEM has evolved* from a space-saving scheme to encode gridded surfaces as quadtrees (its DEPTH component) into a fractal formulation of global geomorphology. Throughout the process, useful properties continued to emerge while the model's size shrank and its simplicity grew. Work remains to be done, especially in regard to the applications (archival, analytic, graphic) that GEM might effectuate or enhance. As well as circumstances have allowed, the origins, procedures and properties of the model have been discussed. Although certain suggestions were made, no particular organization for data was prescribed. This is due both to unanswered questions about the model and to the inevitable tradeoffs involving memory capacity, file access and other properties of particular systems.

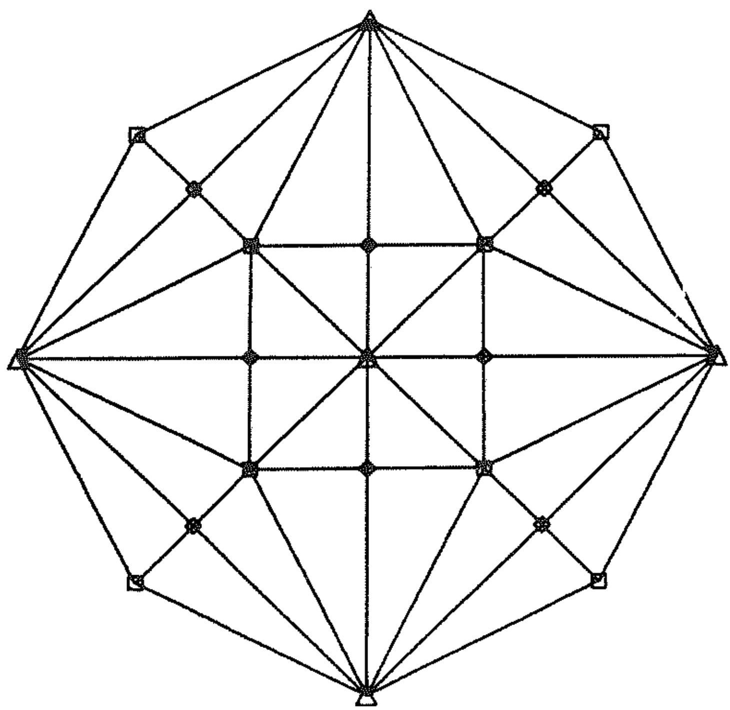
---

*A number of individuals have contributed to this process. Without question, much of its energy devolves from Buckminster Fuller's work, amplified by innumerable other ideas. Among those are Benoit Mandelbrot's elegant expression of elemental eclecticism, les formes fractals, ce n'est pas; while barely hinted at here, this perspective provides important criteria for understanding what the model can contain. Closer to home, Denis White has patiently volleyed ideas off walls where I lobbed them. Dennis Dreher's geometric competence has helped to keep the scheme rooted in physical reality, in the process of constructing a scale model of the data structure. Kelly Chan is gratefully acknowledged for his musings, essays and especially, code. To Dan Schodek, Faculty Director of the Lab for Computer Graphics, go my appreciation for his tolerance and support of visionary puttering. The author, notwithstanding, assumes full responsibility for GEM, in all its polymorphic perversity.

While any given implementation of GEM is likely to be concerned with local terrain data, any of these archives can be merged into larger databases. Their union can be made to contain the most accurate versions of any regions which happen to be duplicated. It is therefore not unrealistic to envision a network of GEM archives, each with detailed data concerning certain localities, functioning as a coherent, reliable but decentralized database.

Claiming that a data model as minimal as GEM can catalog and verify vast numbers of measurements more or less automatically is likely to raise some eyebrows. It is hoped that this presentation has communicated how such properties derive by construction and inference from the polyhedral and polynomial structures serving as the basis for GEM. All persons wishing to comment, contest or collaborate, please contact the author. Together, we may be able to express ideas and information as images of our planet, illuminating its form, features and facets, in all their fullness.

FIGURE 3



Schematic of the initial GEM "trioctacon" lattice: Octahedral vertices are marked with triangles, cubic ones with squares. Intersections of the networks are marked with diamonds. This is an "inverted" perspective, drawn as if viewed from a nearby vanishing point. Note the diamond-shaped facets which connect cubic and octa vertices. These are the faces of a Rhombic Dodecahedron, a form which contains both initial polyhedra.

REFERENCES

Dutton, G. H. (1983) "Efficient Encoding of Gridded Surfaces." Spatial Algorithms for Processing Land Data with a Microcomputer. Cambridge, Ma: Lincoln Institute for Land Policy Monograph Series.

Fuller, R. B., with E. J. Applewhite. (1982) Synergetics: Explorations in the Geometry of Thinking. New York: MacMillan, 876 p.

Fuller, R. B. and S. Sadao. (1982) Spaceship Earth Edition of the Dymaxion Sky-Ocean Map. Philadelphia: Buckminster Fuller Institute.

Fowler, R. J. and J. J. Little. (1979) "Automatic Extraction of Irregular Network Digital Terrain Models." Computer Graphics, vol 13, no 2, pp. 199-207.

Gargantini, I. (1982) "An Effective Way to Represent Quadtrees." Comm. of the ACM. Vol 25, no 12, pp. 905-910.

Gold, C. M. (1978) "The Practical Generation and Use of Geographic Triangular Element Data Structures." Harvard Papers on Geographic Information Systems. Reading, Ma: Addison-Wesley, vol. 5.

Males, R. M. (1978) "ADAPT: A Spatial Data Structure for Use with Planning and Design Models." Harvard Papers on Geographic Information Systems. Reading, Ma: Addison-Wesley, vol 5.

Mandelbrot, B. (1982) The Fractal Geometry of Nature. San Francisco: Freeman, 462 p.

Mark, D. M. (1978) "Concepts of 'Data Structure' for Digital Terrain Models." Proc. of the Digital Terrain Model Symposium. Falls Church, Va: Amer. Soc. Photogrammetry, pp. 24-31.

Peucker, T. K., R. J. Fowler, J. J. Little and D. M. Mark. (1977) Digital Representation of Three-Dimensional Surfaces by Triangulated Irregular Networks. Tech. Rep. 10, ONR Contract no. N00014-75-C-0886, Dept. of Geography, Simon Fraser U., Burnaby, B.C., Canada.

Popko, E. S. (1968) Geodesics. Detroit: University of Detroit Press.

Prey, A. (B. Binder, trans., 1968) "A Mathematical Representation of the Altitude Relationships on the Surfaces of the Earth, Using Spherical Functions of the 16th Order." Harvard Papers in Theoretical Geography. No 22. Originally published in German, 1922.

A STRUCTURED EXPERT SYSTEM FOR CARTOGRAPHY BASED ON THE HBDS

François Bouillé
Laboratoire d'Informatique des Sciences de la Terre
Boîte 1000
Université Pierre et Marie Curie
4, Place Jussieu
75230 Paris cédex 05
France

ABSTRACT

The Structured Expert System, SES, is built around the HBDS kernel. The HBDS (Hypergraph-Based Data Structure) is a knowledge structure which is able to represent any phenomenon of the real world, whatever the complexity may be, both graphically for the user and physically inside the computer. It ensures its own storage and retrieval in a sophisticated data base which is completely portable (among other properties). The expert system thus developped handles rules and facts, which are immediately structured when captured. Owing to the HBDS capability of self-reconfiguration, they can be reorganized at any time, according to some inductive and deductive mechanisms of the motor, thus trying to represent the cartographer's mind. Any research is considerably improved, because it does not depend on the number of rules and facts stored in the system; moreover, HBDS simultaneously supporting both expert system and a very large data base, the system is illimited in size. Using an interface in natural language, it shows that the trend of developping specific programming languages devoted to cartography is now obsolete.

INTRODUCTION

Information science and computer technology applied to cartography had many developments since these last twenty years. Among the most important large trends, we may find the cartographic data bases, the design of specific cartographic languages, and more recently the expert systems applied to different scientific and non-scientific fields.
Are these systems convenient, and are they compatible?
Is the information portable, namely computer-independent?
Is it only stored, or can it be easily used, not only for drawing a map but for some inductive/deductive processes, in an intelligent manner?
In most of cases, the cartographic data bases, or so-called, are nothing but complex file systems, generaly not portable between different computers, with organizations which are not compatible at all, now showing some need of standardization... These systems either have been considered initialy as the storing of some captured informations and there are thousands of magnetic tapes, or they have been developped in accordance with a specific goal, and consequently they cannot be used for another purpose...
Some new programming languages oriented to the cartography have been tried, showing the unadequacy of the low-level languages currently used by the cartographers, such as Fortran, Cobol, or PL/1. It was for compensing the lack, but the result generaly looks like a repaired Fortran or a specific assembly language from many years ago... A better solution could be found in very high-level languages such as SIMULA 67, ADA, or in languages currently used in Artificial Intelligence, namely LISP or PROLOG...But what we now need is the knowledge processing; Kitagawa(1981).

Nowadays, cartographers must be abble to work in natural language, to express the nature and structure of the cartographic phenomenons not only with integer or float concepts, with array or file organizations, but with the set theory components; cartographers must apply some mechanisms like induction and deduction, and that is exactly what an Expert System is abble to offer them.

THE CARTOGRAPHIC PROGRAMMING LANGUAGE : AN OBSOLETE APPROACH

We may distinguish two methods rather different:

An Aborted Trend

Some authors have tryed to develop some very particular languages, specificaly for the cartography. This is not based on a self-extensible language, and the corresponding compiler is generaly written in a very low-level programming language. Because it is impossible at the step of designing a system to take into account all the possible concepts with an exhaustive manner, such languages become immediately insufficiant and moreover they cannot be extended a posteriori.
These languages afforded nothing, and we mention them here only for historical reasons, and for emphasizing how some ways possibly promising considered twenty years ago now appear erroneous.

Another way consists of using a very high-level language (VHLL) allowing abstract data type (ADT) handling.

The Self-Extensible Approach using ADT in VHLL

Instead of digital (or character) data, we firstly need ADT handling, for concepts such as ISOLINE, TOP, SADDLE, BASIN, BOUNDARY, AREA, CITY, RIVER, ROAD, BLOCK, BLOCK_FACE, NETWORK, etc.
Such concepts could be included in a particular programming language, for instance as new extensions in SIMULA 67 or ADA; according to the self-extensibility of SIMULA, we may develop new classes corresponding to the concepts indicated above. Such a way has been used by the author many years ago in SIMULA, and we think that it is very easy to implement in ADA (Bouillé, 1981). But the main problem comes from the fact that all the concepts must be previously defined and cannot be introduced at the executing time. For very simple and limited applications, the self-extensible approach is sufficient. It cannot be used when dealing with large applications with new concepts progressively discovered. Moreover, no deduction can be performed on the data captured. This is nothing more than a VHLL including ADT facilities and accepting a set of new types requiring to be compiled.

> Such an application is self-extensible, may be interactive,
> but the self-extensibility is <u>not</u> interactive.

THE CONCEPT OF EXPERT SYSTEM

Principle

We give here a very brief summary of what an Expert System (ES) is composed of: a knowledge base and a motor.
The knowledge base is made of : a dictionnary, a base of rules, a base of facts.
The dictionnary contains informations on the facts, their definition and their domain.
The base of rules contains the set of rules which can be applied to the facts, and the transformations applied to the rules.
The base of facts contains all the facts, including both those introduced at the initialization of the ES, and those progressively "learned".

The motor is the algorithm corresponding to the "cognitive process"; it allows some inductions and deductions. It can work on logical proposals and on predicates using variables.
For further details on the ES, the Reader may consult for instance Duda and Gasehing(1981), or Laurière(1982). Some interesting applications may be found in Gaschnig(1980),or Clancey and Letsinger(1981).

Advantages

Two main points may be emphasized:

- .with a limited set of restrictions, the user can directly handle the specific concepts of his application in a natural language, with his terminology progressively introduced at running time.
- .the ES is abble to make some deductions, to discover some inconsistencies, to learn new things, and even to deduc new rules.

Problems

The system used to store and retrieve rules and facts is not exactly what we generaly name a "data base". Even, in some ES, it is simply a list of strings loaded in core memory and residing there during the time the ES is runned. The amount of data to be processed is thus strictly limited and the size of the application is considerably restricted. Presently, some researchers try to connect an ES and a real-size data base system, for instance a relational data base. Unhappily, the relational model is not concerned at all by the semantics, and the possible connection is not evident.
An ES is said operational when it runs on a given example; this one is generaly perfectly satisfying for its author, though it generaly is a simple application only chosen for testing the system. Neverhteless, when reaching the size of 1000 rules, the efficiency is often decreasing. It is mainly due to the internal mechanism (the motor) based on a list-processing, with some tree-structures. They are perfectly convenient in compilation technics; they are not the most well-suited structures for handling numerous facts and rules.
We will briefly sum up in the next paragraphs how we avoid these two handicaps.

REAL WORLD MODELLING WITH THE HBDS

The Hypergraph-Based Data Structure (HBDS) is a topological data structure, firstly concerned by the semantical aspect of the phenomenon before any implementation consideration; its main purpose is the representation, banking, and handling of the knowledge. Thus, its goal is not far from that of any ES. Though it has not been developped especially for the cartography, and is presently applied in various fields scientific and non-scientific, it perfectly fits to the complex aspects of the cartography and has found there an interesting field for its progressive experimentation.
The HBDS is based on the set theory and the abstract data type concept. Any phenomenon is made of sets which have elements, which have properties and may possibly present some relationships, either with elements of the same set, or with elements of another set. This basical definition of the set theory is convenient for any phenomenon, among which the cartography.
In HBDS, we have sets named classes, which elements are named objects; the classes have attributes, with kinds and types, and the objects have attributes which are the corresponding values for a given property.
The potential relations are carried by links between classes; when a relation is effectively verified by the objects, these are linked by a "link between objects". We thus distinguish what is potential and what is effective.

There is a graphical representation associated to the different types; a class is associated with an edge of hypergraph; an object with a vertex; attributes are valuations; links are represented by arcs. Moreover a class may embed other classes;and a relation may represent a set of relations; the model is thus self-extensible, because we can recursively combine the four basical concepts of the set theory. The graphical representation (see figure 1) gives the human, cartographers and computer-men, even from different countries, an easy way of communication, and a concrete idea of what they have to deal with, better than with a long text. More developped applications to cartography can be found in Le Graverend (1980), Saadé (1982), Satharanond (1982). But let us show now this simple example.

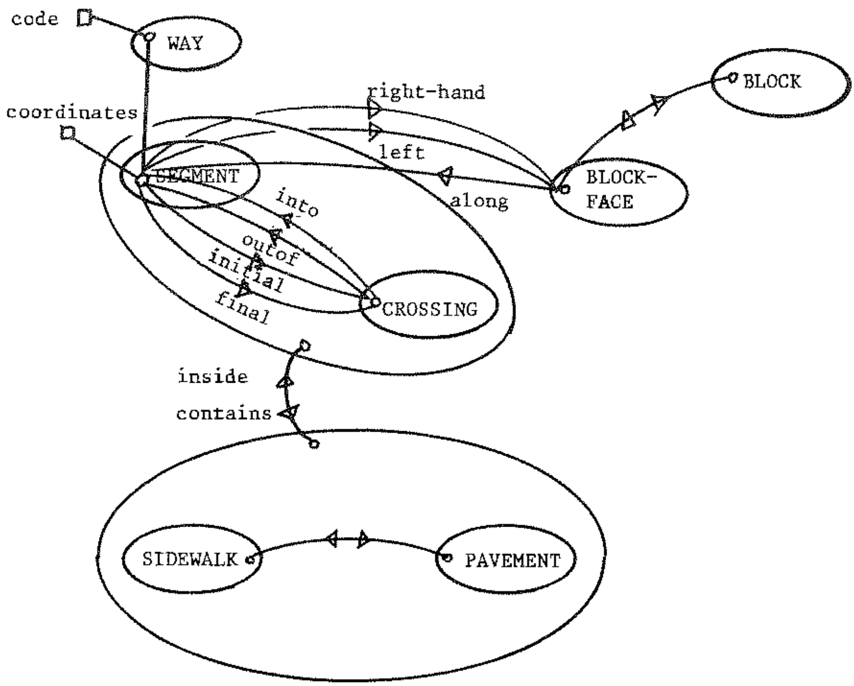
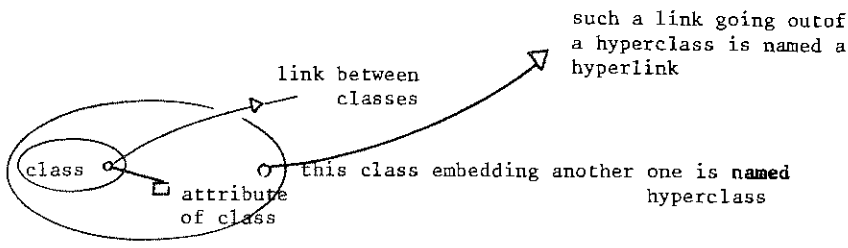


Figure 1 - A very simple example of knowledge structure presenting elementary components of the urban texture.
No object, attribut of object or link between objects have been represented here.



All the elements which have been represented on the picture may be interactively generated at running time, checked, modified, etc.
For the attributes, type, kind, dimensions, etc, can be progressively defined and updated when necessary.
Moreover, any component,among the six ADT previously mentionned, may be doted with two very important characteristics:

-a condition,
-an action.

Condition

Any datum may be conditionned; if a condition has been specified on the datum, each time it is contacted, this condition is tested; it consists of a list of boolean expressions, containing constants as well as variables, including ADT of any type.
The condition may be checked, modified, deleted at running time; it appears as a string, which syntax is in conformity with a list of boolean expressions.

Action

Any datum may be associated with an action which is performed each time the datum is contacted; this action can be compared to a procedure which is activated without having to be called; this action contains any algorythm, whatever complex it may be, working on any type.
This action may consist of some checking followed by some decisions.

These two components, condition and action, make the HBDS very convenient for implementing an ES.
HBDS manages its own data base, with data portable between different types of computers. We do not detail here in this paper dealing with the cartography how it is obtained; we only mention the fact, that an HBDS data base is simultaneously shareable between concurrent processes, protected, portable and distributed. For further details on the HBDS system, the Reader may consult Bouillé (1978, 1979), or some extensions in Perrot(1983).
Another important capability of the HBDS is the automated reconfiguration which is possible without having to modify the whole data base. For for instance, at any time, it is possible to change an object into a class, or an attribute into a link, or to build more complex transformations Revelat (1983).
Combined with the condition and action, the transformation provides the self-reconfiguration capability, and that is exactly what we need.

SES, THE STRUCTURED EXPERT SYSTEM

We have previously seen that an ES needs a dictionnary, a base of rules, a base of facts and a motor. What is directly provided by the HBDS ?
HBDS retrieval is based on a dictionnary which is computer-independant, may be stored in any file system, with its own addressing system. Thus, we do have the dictionnary.
The facts are considered as HBDS basical ADT : classes, objects, attributes (of classes or of objects respectively), links (respectively between classes and between objects); these facts can be restructured when required at any time, owing to the transformation mechanism. These facts are stored in an unlimited data base which is fully portable between any computers. Thus we have the base of facts...
What about the rules?
The rules are immediately represented by the conditions and the actions (when decisions are to be taken). The originality of the HBDS implementation of an ES is in the fact that these rules are directly carried by the components they concern, stored with them in the same data base, and are provided with the same propertis of portability, etc.

Thus we have the base of rules, together with the base of facts, and the dictionnary. Now, we only need the motor of inferency; this one does not work on tree-structure exploration, because we have few trees in the HBDS applications (and moreover, we are convinced that tree structure do not exist in the real world; there are only simplifications introduced by human being before other components will be available...). What is then the general structure of the SES ? The figure 2 sums up, very simplified, the different concentrical abstractions of the HBDS system, with a supplementary layer for the SES.

Why is it named Structured Expert System? Because everything is structured inside; dictionnary, itself, facts and rules; this last point being the most carcateristic, ensuring a strict independency between time processing and the size of a large application; when accessing some data, you directly access the only rules concerning them, and all of them; these rules in fact have been immediately dispatched when introduced at their convenient place into the knowledge structure.

Another particular point is about the motor. This one can be changed without the whole structure (rules, fact and dictionnary) is modified. Likewise, the layer dealing with the natural language may be independently replaced. These points are particularly important for the progressive evolution of a large application during a lot of years.

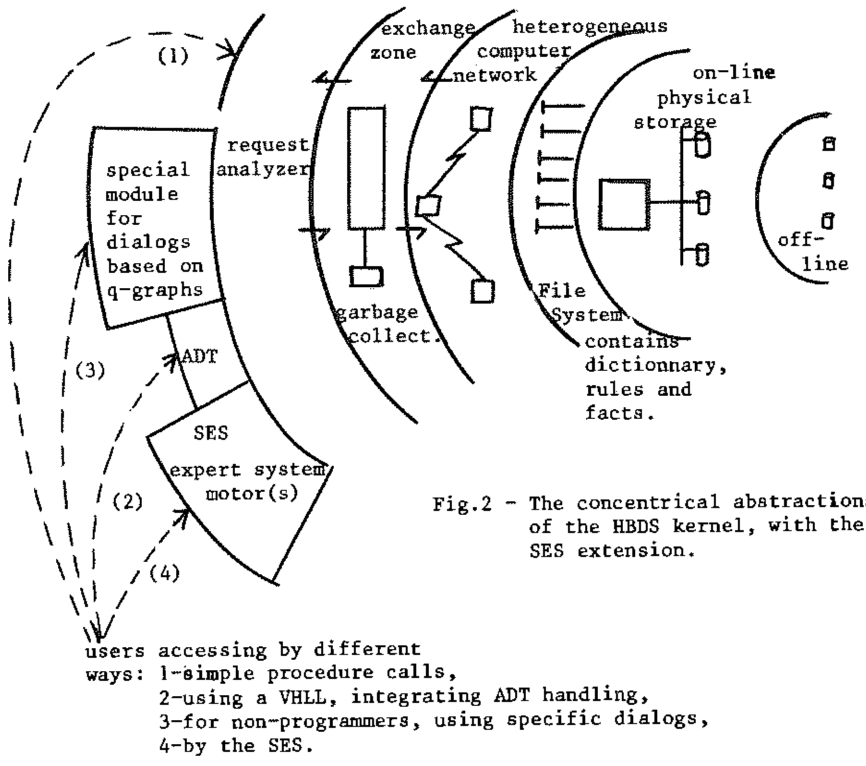


Fig.2 - The concentrical abstractions of the HBDS kernel, with the SES extension.

Though the distribution is not the main topic of our paper, we must emphasize at that point another original aspect of the SES; instead of one only motor for the whole structure, we may consider several motors, possibly different, carried by the data structure, like rules and facts.

We now may present some very simple examples dealing essentially with the cartography, in order to give a better idea of what an ES looks like.

A CARTOGRAPHIC SES : SOME SIMPLE EXAMPLES

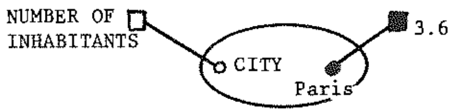
We briefly present some assertions in natural language, and the corresponding data structure which is generated and stored into the data base. Example :

PARIS IS A CITY

A CITY HAS A NUMBER OF INHABITANTS EXPRESSED IN MILLIONS

THE NUMBER OF INHABITANTS OF PARIS IS 3.6

The result is shown on the figure 3 below.



Second example :

A WAY IS COMPOSED OF SEGMENTS

A SEGMENT HAS A LEFT_HAND SIDE BLOCK_FACE AND A RIGHT_HAND SIDE BLOCK_FACE

A WAY HAS A CODED NUMBER

A SEGMENT HAS AN INITIAL CROSSING AND A FINAL CROSSING

A CROSSING HAS COORDINATES

A SEGMENT MAY CONTAIN A PAVEMENT AND ONE OR SEVERAL SIDE_WALKS

A SEGMENT HAS AUXILLIARY COORDINATES

etc...

The Reader has probably recognized a part of the structure which is shown on the figure 1; in this example, we see that new types have been introduced at run-time, for instance : WAY, SEGMENT, etc, without a special programming language.
Naturally, we can add some mechanisms, for instance rejecting a proposal such as : a segment is composed of ways, which will activate an inquirement to the user, because it is not compatible with the proposal : a way is composed of segments...

USING OR NOT AN EXPERT SYSTEM

Some applications do not require the use of an expert system, such as systematical data capture. One of the main advantages of the HBDS/SES is in the possibility of using HBDS without the ES, the data being stored exactly as if they were processed through the SES.
We must avoir a systematization in the use of the ES. Classical programming is perfectly convenient for a large area of applications; HBDS can be contacted with some procedure calls, or by using some abstract data types facilities which exist in a programming language especially designed for HBDS and which name is precisely ADT'81. For most of routine works, we do not need an ES. But for many new applications, or for specify

ing a new knowledge structure and be sure it is correctly expressed, then we must orient our mind and habits to this actual tool. In any cases, the ES must not be considered as a stand alone tool, but it must be integrated into the complete cartographic system, in a very good place corresponding to the potentiality of the tool for the cartographers, possibly in connection with other professional areas.

CONCLUSION

The Expert Systems represent a new family of tools from information science applied in cartography; it is not separated from other methodologies, and can be connected to data bases, and other technics; it is only a new branch allowing to introduce a computer intelligent processing in the cartography to day, by induction and deduction mechanisms.
Such an ES must be optionnally proposed in a cartographic system, but must not become an obligation when not necessary; it does not delete the classical processing methods; it replace them for the sophisticated cases but must be compatible with them, for combining different steps, and by the data which are stored and exchanged between the steps in time, and between the companies in space.
An ES is not limited to a simple set of data; it must be applied to large and concrete and complex cases with a great number of rules, thus requiring to refine the structuring methodology. That is why the SES has been designed and finds in cartography an interesting field of investigation, with a reciprocal benefit.

REFERENCES

Bouillé, F. 1979, The HBDS and its Applications to Data Structuring and Complex Systems Modelling, Report to the Air Force European Office of Aerospace end Development, 167p.

Bouillé, F. 1979, The Use of Hypergraphs in Data Structures Applied to Cartography, proceedings of the Auto-Carto IV Conference, 10p.

Bouillé, F. 1981, The Impact of ADA on Cartography and Topological Terrain Modelling, proceedings of the ASP-ACSM Fall Technical Meeting, 8p.

Clancey, W., Letsinger, R. 1981, Neomycin, Reconfiguring a Rule-Based Expert System for Application to Teaching, IJCAI, vol.7, pp.829-836.

Duda, R., Gasehing, J. 1981, Knowledge-Based Expert System Come of Age, RAIRO, vol.15, pp.325-341.

Gaschnig, J. 1980, Prospector, Development of Uranium Exploration Models for Prospector, SRI Report n°7856.

Kitagawa, T. 1981, Datalogy and Semiology in Establishing Knowledge-Information Processing Systems in Research, proceedings of 2nd Int.HBDS Sem., pp. 1-24.

Kitagawa, T. 1981, Generalized Relative Ecospheres of Knowledge-Information Processing Systems and their Implications for Scientific Research, proceedings 2nd Int. HBDS Sem., pp. 25-42.

Laurière, J.L. 1982, Représentation et Utilisation des Connaissances: Technique et Science Informatiques, vol.1, n°1,2.

Le Graverend, F. 1980, Cartography and Urban Planning: a Tool for Decision-Makers, proceedings of HBDS W.G.on Topological Data Structures and W.G. on

Topological Terrain Models at the 10th I.C.A. Tokyo Conference, 8p.

Perrot, Y. 1983, Modèlisation et Simulation d'un Banque de Connaissances à Exécution segmentée, Parallèle et Itinérante, Thèse de Doctorat d'Etat, Univ.Curie, 600p.

Revelat, J. 1983, Les Transformations Automatiques de Structures, Thèse de Doctorat d'Etat, Univ. Curie, 150p.

Saadé, M. 1982, Structuration de Processus et Simulation, Application aux Réseaux de Transport Urbains, Thèse de 3ème c., Univ. Curie, 160p.

Satharanond, V. 1982, Méthodologie de Structuration des Connaissances et de Saisie du Répertoire Géographique Urbain, Implémentation et Mise en Oeuvre, Thèse de 3ème c., 180p.

BUILDING A HYPERGRAPH-BASED DATA STRUCTURE:

The Examples of Census Geography and the Road System

by
Robert D. Rugg
Department of Urban Studies and Planning
Virginia Commonwealth University
Richmond, Virginia 23284-0001
U.S.A.

Abstract

To illustrate the problem of data structuring, two distinct kinds of spatial phenomena are examined, and their interrelationships explored. These two are the road system, commonly viewed in the context of a network model, and Census geographic units which are commonly represented as a nested spatial hierarchy. Through this exercise, using Bouillé's Hypergraph-Based Data Structure methodology, essential characteristics of each structure will be identified and several key concepts such as "hyperclass" and "multilink" will be elucidated. It will be shown that (1) the network model provides an incomplete view of the structure of the road system; (2) the concept of spatial hierarchy provides an incomplete view of the structure of Census geographic units; and (3) the HBDS model provides not only a more complete view of each phenomenon separately, but also a basis for specifying the relationships between two phenomena with distinctly different inherent structures.

Using HBDS, not only is a fuller, more complete view of each phenomenon obtained, but the two are readily incorporated in the same data structure model. Through such higher level constructs as "hypergraph," "hyperlink," and "multilink," a generalized overview of the complete structure is obtained without sacrificing any of the detailed information concerning specific relationships of individual objects within the larger system.

Introductory Concepts

Hypergraph-Based Data Structures are built on four fundamental concepts, which taken together are sufficient to describe any real world phenomenon: object, class, attribute, and relation. The concepts combine to form the six basic "abstract data types" of the HBDS model: object, class, attribute of object, attribute of class, relation between objects, relation between classes. Figure 1 illustrates the symbolic representation of these data types as developed by Bouillé. "Objects" of "Classes" correspond to "elements" of "sets" in set theory; an "Object" may also be viewed as a vertex in graph theory; an "Attribute" is a property of an object (viewed as an element or as a vertex);

"Relations" correspond to links (either "arcs" or "edges") in graph theory. A hypergraph (Berge, 1970) is a graph in which more than two vertices are linked by the same edge, hence allowing for the manipulation of "Classes" as well as "Objects." The theory of hypergraphs thus provides a conceptual basis for combining set theory and graph theory in the same data structure model. (The complete model is presented in Bouillé, 1977.)

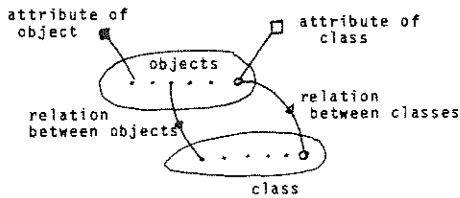


Fig. 1 - Six Abstract Data Types of the H.B.D.S. Model

Census Geographic Units

Census geographic units include (among others) states, cities, counties, towns, minor civil divisions, tracts, block groups and blocks. These fit more or less well into a hierarchy of areal units, as blocks nest into block groups, block groups into tracts, tracts into cities and counties, and the latter into states. A simplified view of the structure of Census geographic units is presented in Figure 2.

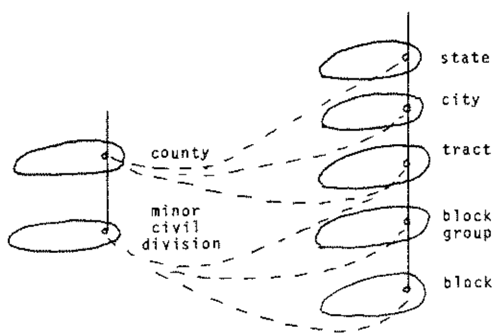


Fig. 2 - The "Hierarchy" of Census Geographic Areas

The circles indicate types of geographic areas, and the vertical lines indicate a hierarchical dependence among the various classes. (The relationships could also be diagrammed as a rooted tree, with the State as the root.) However, the simplified view is already more complex than a hierarchy in the ordinary sense. Two hierarchies are present: in Virginia for example, cities are geographically and juridically independent of their adjacent counties. The relationships between the two "trees" also pose several complications. Both trees have the same root, the State; in Virginia, cities are "county equivalents" for some purposes but not others (e.g., counties are subdivided into minor civil divisions whereas cities are not); portions of some counties are tracted while others are not, and the tracts and block groups in the portions of counties which are tracted may be split by, rather than nesting into minor civil divisions. Finally, again using the example of Virginia, the Census Bureau recognizes three kinds of "places": independent cities, incorporated towns, and "census designated places" or unincorporated towns. Places, also, may or may not nest within a given magisterial district. (See Figure 3.)

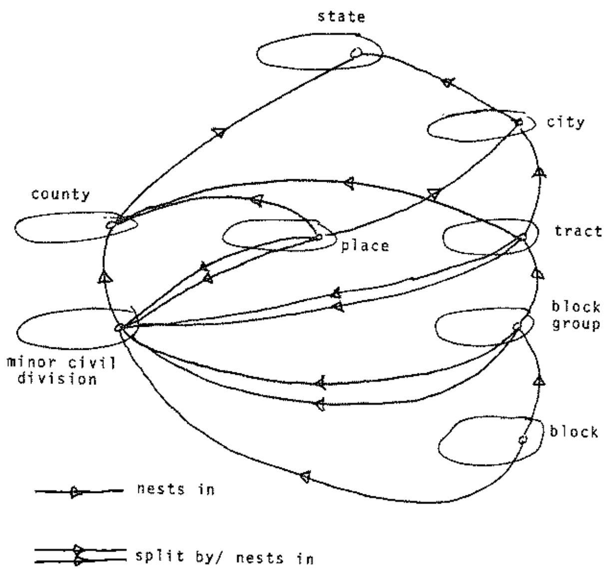


Fig. 3 - A More Complete View of Relationships Among Census Geographic Areas

Now, considering the full range of relationships indicated in Figure 3, it is clear that the hierarchical model is inadequate to fully describe the relationships that actually exist. A next step in HBDS methodology is to find commonalities in some of these relationships and, in so doing, to simplify the structure to present a clear, more generalized view of the phenomenon. In Virginia, where cities are independent of counties, counties and cities have many equivalent relationships and may thus be treated, for some purposes, as a "hyperclass:" a set of classes whose objects may each verify the same relationships with objects of other classes in the system. Counties and cities share the same relationships with the state as a whole and with tracts and places (viewing the independent city as both a "place" and as the county equivalent it nests within, in accordance with Census Bureau usage). Counties and cities thus form a "hyperclass" and the relationships of inclusion between them and certain higher and lower order geographic units are "hyperlinks."

The twin relationships "split by" or "nests in" apply to tracts, block groups and places in relation to minor civil divisions. Tracts and places thus form a hyperclass in their relationships to cities and counties on one hand, and minor civil divisions on the other hand. The relationships "split by" or "nests in" can be subsumed under the HBDS concept of "multilink", thereby further simplifying the diagram. The multilink between minor civil divisions and tracts and places, the latter being a hyperclass, would be termed a "hypermultilink." The power of the HBDS model can begin to be understood when one realizes that, with relatively little effort, the 14 relationships indicated in Figure 3 have now been reduced to only 7, as shown in Figure 4. It is important to recognize that, while the diagram has been simplified, none of the information concerning the relationships among the various geographic units has been lost in the process.

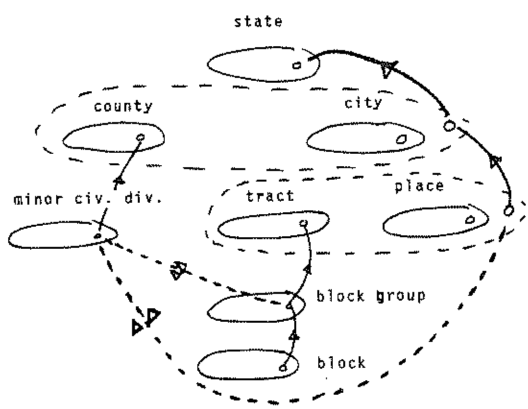


Fig. 4 - Census Geographic Area Struture Using Hyperclasses and Multlinks

Relationship to the Road Network

The road network relates to Census geography in that it forms the boundaries of Census units. However, not all Census areas are bounded by roads: rivers, railroad lines, utility rights-of-way, invisible administrative boundaries, and other features are also used. Bouillé (1980: Fig. 9-11) has already solved the analogous problem of relating administrative entities to administrative boundaries. Following his initial model, all of the Census geographic units shown in Figure 4 can be considered a hyperclass, related to a second hyperclass of Census area boundaries. (These are defined as hyperclasses because, for example, a census block boundary may also be part of a county boundary, and in general all types of boundary lines potentially help define the limits of all types of Census geographic units. See Figure 5.)

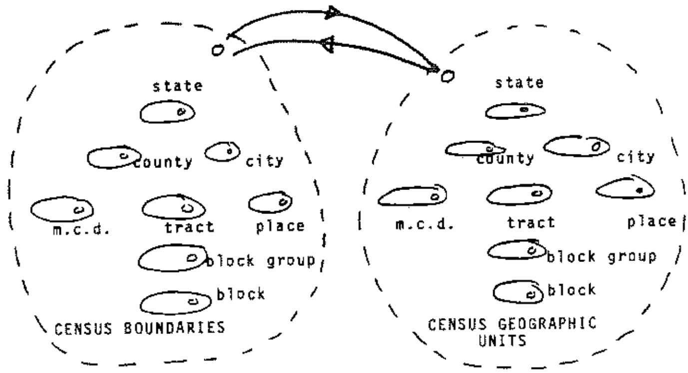


Fig. 5 - Census Geographic Areas and Boundaries (after Bouillé)

Considering that the road network is but one element of the various features used to define Census boundaries, the problem of relating Census geography to the road network can easily be solved by introducing a third hyperclass of features from which the various kinds of Census boundary lines are selected, as shown in Figure 6. However, this view does not take into account the structure of the road network itself, a problem which was left unresolved at the outset of this discussion.

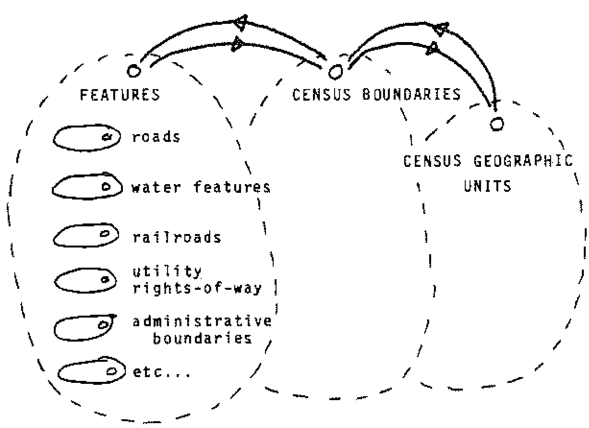


Fig. 6 - Census Areas, Boundaries, and Boundary Features

The Road System

The "road system" consists of various types of roads (interstate highways, U.S. routes, state highways, county roads, etc.). Although it is possible to conceive of each road type as a separate class, in general it is sufficient to consider the type of road as an attribute of the class "road." The class road is illustrated in Figure 7 as the class "whole roads," in order to emphasize the distinction between whole roads as features and road segments as elements of a conventional transportation network, described below.

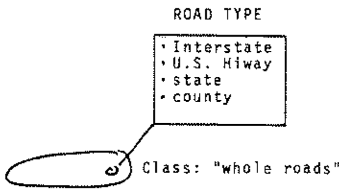


Fig. 7 - Whole Roads and Road Types

In a classical network model, emphasis is given to the intersections between whole roads (nodes, or vertices, of the road network), and to the resulting road segments connecting each pair of intersections (links, or in graph terminology, "arcs" or "edges" of the road network graph).

A complete view of the road system incorporates both perspectives: roads as features and roads as the building blocks of a transportation network composed of nodes and links. Roads, intersecting with each other, produce road segments which, together with the intersections, form the graph of the road network. Further, the road segments are related to the whole roads of which they are a part, sharing certain of their attributes including the type of road. The road system can thus be fully described in terms of the relationships among three classes-- whole roads, intersections, and road segments--, together with the attributes of each class (see Figure 8). Of these three, it is the class "road segments" that forms portions of the boundaries of Census areas. As it turns out, the result presented in Figure 8 is the same for any network, as has been demonstrated in LeForestier (1979) and LeGraverend (née LeForestier) (1980). It depends on transforming one abstract data type (intersection, which begins as a loop on the class "whole roads"), into another (the class intersections, which constitute vertices in the graph of the road network, and divide whole roads into segments, the latter constituting the arcs of the road network graph).

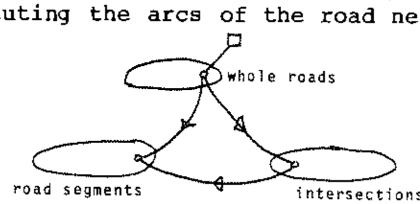


Fig. 8 - Structure of the Road System

Census Geography and the Road System

If all Census geographic areas were bounded by roads, the problem of relating Census geography to the road network could be solved at this point by a simple extension. Regardless of the higher order structure of Census geographic areas, the fundamental unit is the Census Block, which nests into all higher level areas and is never split by any of them. Beginning with the graph of the road network (say, graph G), one can define the dual (G*) as the graph of areas bounded by roads and intersections (each such area represented as a vertex of G*), having links of adjacency defined as the boundaries between the areas (the arcs of graph G become arcs of G*). Each Census block would then be composed of one or more of the areas defined in G*. The simplest case is illustrated in Figure 9. Census Block 101 is bounded on the north by Prospect Avenue, on the east by Second Street, on the south by Glendale Avenue, and on the west by First Street. The physical block defined by the road network as block (1) is the same as Census Block 101.

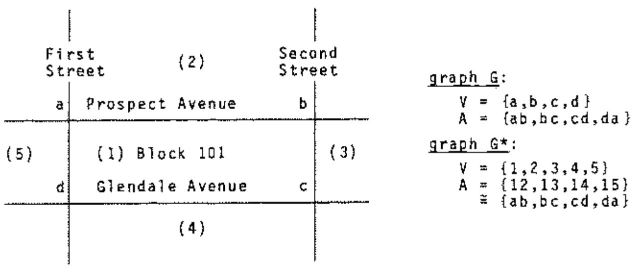


Fig. 9 - A Census Block Bounded by Roads

The definition of areas bounded by the road network results in a new class "areas" and the extended general solution to networks of all kinds, including the areas bounded by network segments, presented by Bouillé (1983), as shown in Figure 10.

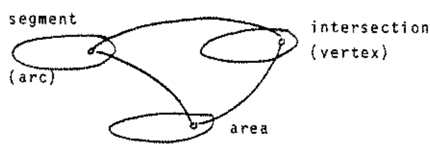


Fig. 10 - Generalized Network Model (after Bouillé)

Unfortunately, the reality of Census boundaries is more complex, involving not only roads but many other kinds of features that potentially serve as the boundaries of Census Blocks. Consider hypothetical Census Block 102 (Figure 11). It is bounded on the north by Prospect Avenue, on the east by a Pennsylvania Power and Light Company power line, on the south by Glendale Avenue, and on the west by Second Street. It is intersected by Harry's Brook, which is not a census boundary here, but is used as a census boundary elsewhere along its course. To define the area of Block 102, it is necessary to consider all the intersections between the road network, the electric power distribution network, and the stream network. Each of these networks, of course, has its own structure, in general composed of classes analogous to those of "whole roads," "road segments," "intersections," and "areas bounded by roads." However, for our immediate purpose, the structure of each network considered separately is not important; rather, we are interested in a different network, composed of all the segments and intersections formed by all the features that potentially serve as Census area boundaries.

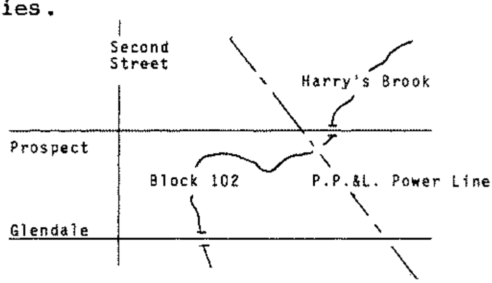


Fig. 11 - Hypothetical Census Block 102

The more general "feature network" consisting of the segments formed by the intersection of all such features is illustrated in Figure 12. The segments defined by the intersections of all features become the building blocks of

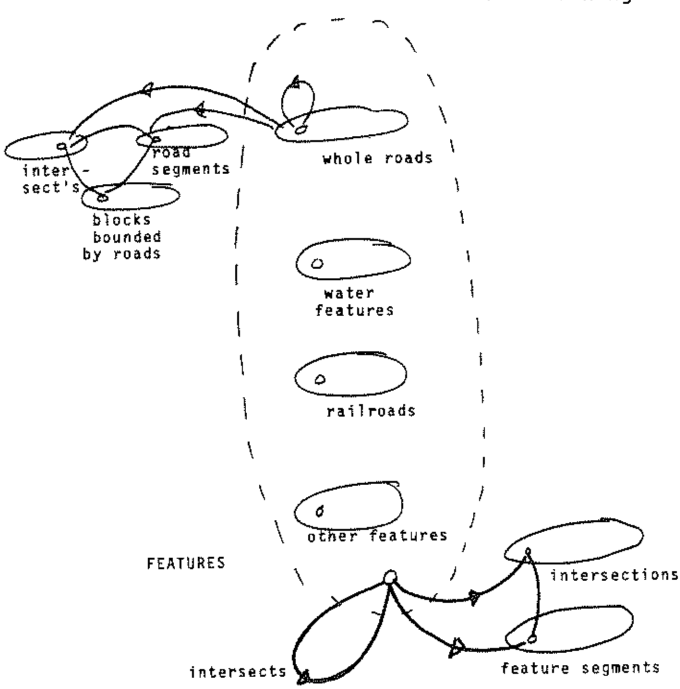


Fig. 12 - Generalized Feature Network

Census area boundaries. The complete structure, relating Census geographic units to each other and to the road system, is presented in Figure 13.

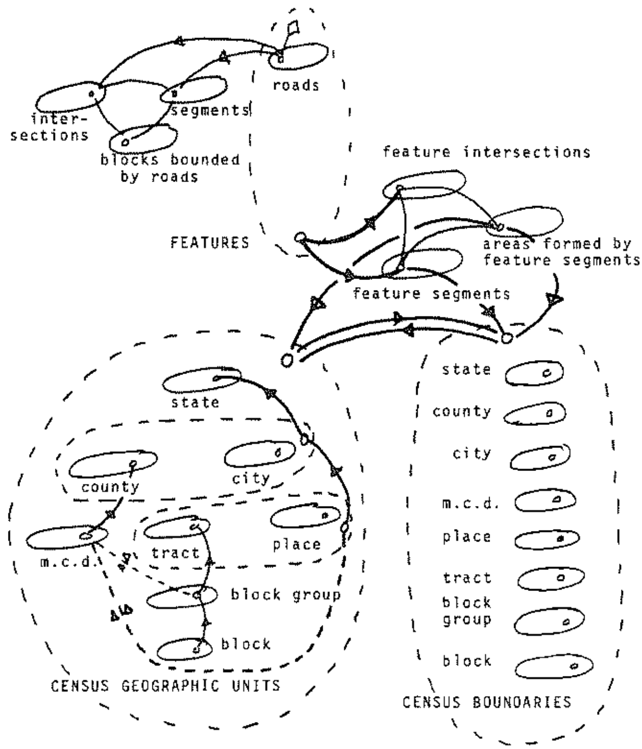


Fig. 13 - Census Geography and the Road System

Conclusion

In describing the structure of the road system, obviously some information about the system is lost if it is modelled simply as a network. Conceiving of "whole roads" as entities within a hierarchical model does not fill the gap, either. The notion of hierarchy, while tempting to apply to Census geographic units, fails to describe all of the potential relationships among elements of the latter system. Without using HBDS methodology, it would be all but impossible to include such divergent phenomena as a road network and the quasi-hierarchy of Census geographic areas within the same data structure model. Using HBDS, not only is this task made possible, but also, through such higher level constructs as "hypergraph," "hyperlink" and "multilink," a generalized overview of the structure is provided, without loss of any of the detailed information concerning specific relationships of individual objects within the larger system.

REFERENCES

Beattie, A.D. and R.D. Rugg (1982). Census User Workshop Manual. Richmond: Virginia Commonwealth University, Center for Public Affairs, mimeo.

Berge, C. (1970). Graphes et Hypergraphes. Paris: Dunod. English version: Graphs and Hypergraphs, translated by E. Minieka, New York, North Holland Publishing Company, 1973.

Bouillé, F. (1977). Un Modèle Universel de Banque de Données, Simultanément Partageable, Portable et Répartie. Thèse de Doctorat d'Etat ès Sciences Mathématiques (mention Informatique), Université Pierre et Marie Curie.

__________ (1980). Contributions of Graphs and Hypergraphs to Cartographic Information Theory. Translated by R.D. Rugg from "Apport des Graphes et Hypergraphes en Informatique Cartographique", Bulletin du Comité Français de Cartographie, March 1977. Richmond: Virginia Commonwealth University, Center for Public Affairs.

__________ (1983). Lectures presented in a week-long series at HBDS Tutorial '83, Richmond, Virginia.

LeForestier, F. (1979). Les Structures d'une Banque de Données Urbaines. Proceedings of the First International "HBDS" Seminar, Lisbon, Portugal.

LeGraverend, F. (née LeForestier) (1980). Cartography and Urban Planning: a tool for decision-makers. Communications Presented at the Tenth International Conference of the International Cartographic Association, Tokyo, Japan.

MULTIPLE DATA STRUCTURES IN A REGIONAL DATA BASE

Paul M. Wilson
Geogroup Corporation
2437 Durant Avenue
Berkeley, CA 94704

ABSTRACT

Early geographic information systems were limited in the ability to represent spatial data in different ways. Advances in hardware and software have made possible a greater degree of flexibility, allowing multiple data structures to be used in a single application. A higher level of data integration is possible when the limitations on data representation are removed. A regional data base (BASIS in the San Francisco Bay Area) is used to illustrate these concepts. Several BASIS applications, ranging from site-specific analyses to regionwide studies, demonstrate the use of multiple data structures (cells of different sizes, vectors, polygons).

INTRODUCTION

The techniques and practice of geographic data handling have advanced very rapidly since the days of punch cards and line printer maps. Aided by the amazing progress in computer hardware and software, the automated handling of spatial data has become a widely used tool in both government and the private sector. A new acronym, GIS (for Geographic Information System) has come into common use to describe computer mapping and related data acquisition, storage, and manipulation techniques.

The evolution of the GIS field has seen many changes in theory and practice. Some of these changes (such as the ability to handle very large data bases, for example) have resulted almost directly from advances in computer hardware. The availability of minicomputers and microcomputers has lowered the costs of systems and made them more accessible to new users. Other examples of hardware-driven changes include the use of digitizers (and more recently, scanners) as input devices and the proliferation of options for hard copy output.

Other advances are more grounded in software. The overall pattern of software development over the last decade has been from batch to interactive, from "computer room" rigidity to "user-friendly" environments; new software tools have been designed to take advantage of more powerful hardware. Research in data structures has led to the application of methods such as relational data bases and quadtrees to geographic data handling.

Another major factor in the evolution of GIS practice lies in demand. The decade of the 70s saw an increasing attention to environmental issues; advances in the understanding of natural processes and the interactions of man and nature led to a demand for much more data and ways to manage that data. Complex applications require large data bases, and spatial data handling techniques have been forced to change to meet these demands.

Many of these changes arise directly from institutional factors. The nature of applications is dependent on the needs and the resources of users. Early GIS applications dealt with single issues, since frequently the organization developing the system only needed specific products. Fiscal realities have reduced the number of systems that can be built for a single purpose; budget cuts limit the ability to build new systems unless the systems can pay for themselves by performing many functions. These fiscal constraints argue for integrated systems -- systems that can serve multiple application areas and/or multiple users.

THE CHALLENGE OF DATA INTEGRATION

Data integration is a widely used, and often misunderstood, term. There are a number of valid meanings, because integration is possible in several senses. Each type of integration may exist with, or even complement the others. The most common use of the term refers to integration of the data itself, seen from the perspective of a single user. This type of integration results in a data base where many different types of data can be stored and interrelated. The data base can be viewed as a set of data planes or a combination of attributes associated with each spatial unit. Some systems today include natural features, networks, and demographic characteristics. The central characteristic of such a system is the ability to store and manipulate more than one type of data for a given geographic area.

A second way of looking at integration is in terms of applications. Here, a single data set which is collected for one application may be used for others produced by the same GIS. This requires planning in the data collection stage as well as flexible data management, since different applications tend to view the same type of data in widely different ways.

The most complex meaning of integration involves institutions. In this sense, a given data set (or even an entire system) would be accessible to many organizations. It is similar to the applications integration meaning above, but with the complication of multiple organizations (usually with different geographic information systems) thrown in. Anyone with an interest in a specified geographic area would be able to obtain and utilize spatial data collected by other organizations for that area; cities, counties, regional groups, state agencies, and federal agencies could share data in a "vertical" fashion.

Almost everyone agrees that data integration (in any of the above meanings) is good. Its clearest value is in efficiency. Substantial cost savings can result if data collected for one purpose can be used for other purposes. Another, and more subtle, advantage, lies in consistency: common use of a data base removes one of the potential variables that arise in studying complex issues.

If data integration in a GIS is such a universally recognized benefit, why is it so difficult to achieve ? Examples of successful data integration are difficult to find, and those cases which have worked often consumed far more time and expense than had been anticipated. The basic reason for the difficulty in integrating geographic data is that there are a series of conditions which must be satisfied, and the failure to overcome any one of them will prevent integration.

It is easy to identify some of these factors. They include area of coverage, data classification, accuracy, representation, geographic referencing, resolution, volume, temporal factors, and transfer medium. Despite work being done on the definition of exchange standards, there are still significant technical problems to overcome before data sharing is common.

Other factors — the ones which usually prove most difficult to overcome — are institutional. Barriers which arise from organizations or individuals are usually more difficult to identify, and always more difficult to overcome. They include timing, public/private issues, legal constraints, cost factors, political problems, and even personality conflicts.

The focus here will be on technical issues, specifically that of data representation. One of the basic problems in any GIS is that there are many common modes of spatial data representation. The source of any data set usually determines its mode of storage. Digitizing maps results in a vector or chain representation, for example, while scanning them yields a raster file (which can be converted to vectors). Some sources of machine-readable spatial data (e.g., Landsat and digital elevation models) are raster mode in origin. Conversion from one mode to another requires time and money.

A basic question in determining suitablity of representation modes is the nature of the target application. For high-quality graphics, a vector mode is often preferred since it most closely mirrors the way that people are used to seeing maps. On the other hand, overlays and other forms of data manipulation are most straightforward in a cell or raster mode. The issue is made more complicated if some form of data integration is a goal; the same data set may be needed in different forms by different users.

Conceptually, the solution to this dilemma is simple: collect the data in a form which can be readily converted to any necessary structure. In practice, this is often not easy. The use of a flexible scheme for digitizing allows later conversion to cells or polygons, but data which is supplied in raster format cannot be as readily converted.

BASIS AS AN EXAMPLE

BASIS, the Bay Area Spatial Information System, is a good illustration of the ways in which multiple data structures are used in an operational data base. BASIS is a large GIS covering the 7000-square mile San Francisco Bay region (Figure 1). It was created by the Association of Bay Area Governments (ABAG) in 1974, and was designed to support regional and local planning applications. The system is now maintained and operated by Geogroup Corporation, and is used to provide services to both governmental and private organizations.

One of the guiding objectives in the design of BASIS was the ability to support many different applications. ABAG's scope was very broad, ranging from land use planning to environmental management and hazard mitigation. This scope meant that the system would potentially be called on to provide many diverse types of products. The current listing of data contents (Figure 2) is a good indicator of the range of data types that have been needed.

In addition to the number of data sets, data sources must be considered. A regional system, more than a local or statewide one, must rely on the integration of outside data sources. Agencies at the local or state level often conduct their own programs of data gathering to support operational needs. Few agencies with a regional scope have the broad operational responsibilities held by units

FIGURE 1

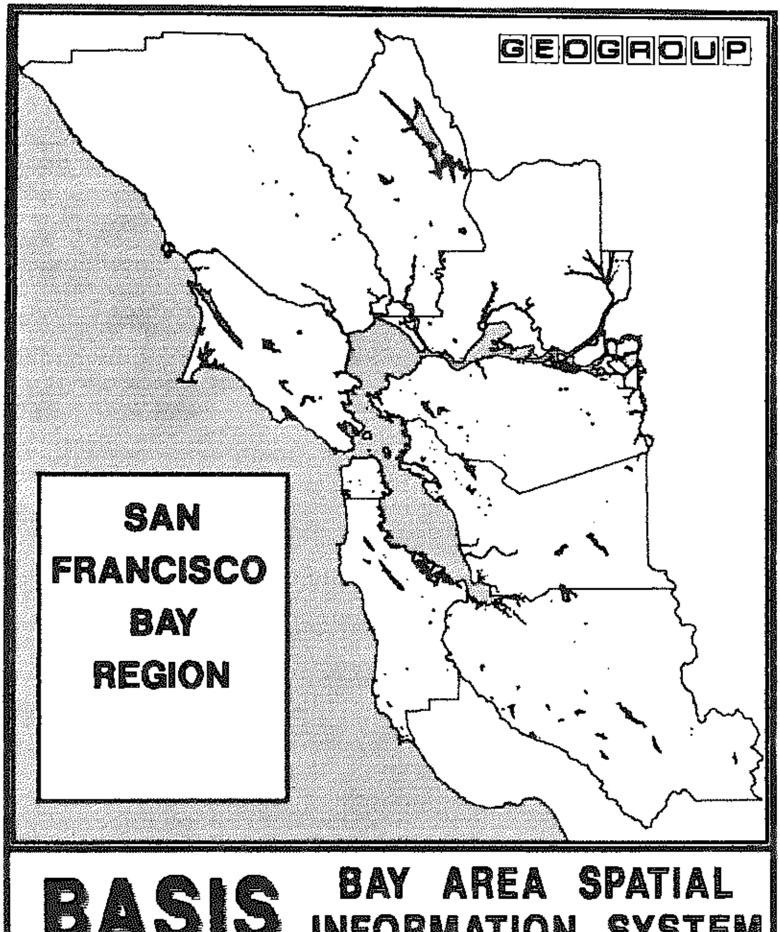

GEOGROUP
SAN FRANCISCO BAY REGION
BASIS
BAY AREA SPATIAL INFORMATION SYSTEM

FIGURE 2: BASIS Data Base Contents

NATURAL ENVIRONMENT
Bay Water Depths
Elevation (Digital Elevation Models and Digital Terrain Tapes)
Geologic Materials
Hydrography (Bay and Ocean Coastline, Lakes, Marshes)
Precipitation
Slope / Aspect
Soil Associations
Wind Speed

POLITICAL / ADMINISTRATIVE UNITS
Census Tracts (1970 and 1980)
City Spheres of Influence (LAFCOs)
County Boundaries
Sewage Districts
Solid Waste Collection Areas
Transportation Zones
Water Districts
ZIP Codes

TRANSPORTATION
Bridges / Overpasses
Electric Power Lines
Gas Pipelines
Highways
Petroleum Pipelines
Street Network (DIME) Files
Water Aqueducts

LAND USE
Airports and Airport Runways
Hazardous Waste Sites
Industrial Sites
Land Use
LANDSAT Land Cover
Regional Parks
Seaports
Sewage Treatment Plants
Urbanized Area Boundaries

HAZARDS
Air Pollutants (six categories)
Airport Noise Zones
Dam Failure Inundation Areas
Earthquake Intensity, Maximum
Earthquake Risk (three building types)
Fault Study Zones and Fault Traces
Flood Plains (FEMA-NFIP and USGS)
Landslide Susceptibility (Earthquake-Induced, Rainfall-Induced)
Landslides
Liquefaction Potential
Liquefaction Susceptibility
Tsunami Inundation Areas

of government. This is certainly true of BASIS: neither ABAG, as a regional planning agency, or Geogroup, a private company, have extensive data collection programs.

The primary unit of data representation in BASIS was originally the one-hectare grid cell (100 meters square). Coverage of the region (nine counties, the Bay, and small areas of the ocean which contain data of interest) requires over two million of these hectare cells. This structure was selected as a compromise between level of detail and ease of implementation. It was clear that a larger cell (perhaps a quarter square kilometer) would support any project that was regional in scope. A smaller cell would allow for more detailed local studies, but would increase the costs of implementing and maintaining the system. Other data structures, such as polygons, seemed to be stronger conceptually but much more complex to implement. The hectare cell was chosen as a structure that maximized the number of potential applications while assuring a good chance of making the system work.

Since ABAG also provided technical assistance to local governments, another system design goal for BASIS was the ability to handle small areas with a greater level of detail. It was clear that the hectare cell structure would be adequate for many subregional studies and a few local applications, but could not support sufficient detail for site-specific projects. There was no immediate method (or need) for attaining the capability for more detail; however, all original data encoding and digitizing was designed to be retained so that conversion to a smaller cell or to other structures could be accomplished later.

This capability to maintain a regional overview, while capturing local detail where appropriate, is very important. It adds the ability to support applications at both regional and local scale, and promotes a consistency of data used for decisionmaking at several levels of government. The impact of these needs on data structures is contradictory, however. One end of the spectrum, the regional perspective, calls for a data base which can cover a large area at a relatively low cost. The other requires detail, and does not care about extensive area coverage.

Both of these factors act to demand flexibility in data structures. Having many different types of data makes it more unlikely that any single data structure will provide a satisfactory representation. The need to vary the level of detail means that no fixed-size unit can be used, and that a mode which is adequate at one scale may be too gross or too detailed at another scale.

BASIS has been utilized for many applications at both the regional and local scale (see References). These include the location of sites for hazardous solid wastes, several studies of earthquake hazards, industrial site data, airport noise analysis, market feasibility studies, and mapping of socioeconomic data. Three of these projects serve as good illustrations of the need to incorporate multiple data structures.

The first, use of the system for local environmental assessment, points out some of the limitations of a cell-based system. This application required the overlay of many types of environmental data with manmade features (property boundaries, transportation networks, land use patterns). Similar procedures at the regional scale had worked well; the overlayed environmental data sets were generalized to hectare cells (usually in a dominant area procedure, although data types such as landslides were coded for the presence of any amount in the cell) and then overlayed.

This method, while adequate for studies at a regional scale, is clearly limited for most local needs. A city planning agency will usually need data concerning parcel-specific land use as part of its analysis. This level of detail is difficult to capture or maintain in regional data bases. Trying to generalize the parcel data into large cells is unsatisfactory; on the other hand, collecting environmental data in terms of parcels is a costly and time-consuming process. The ideal solution is to represent each type of data in the mode which is most appropriate to that type, and develop methods to derive the spatial relationships among each.

In practice, this results in data stored in several different modes of representation. The parcel is stored as a polygon. Environmental data is stored as cells, with cell size being determined by the source and quality of each data item.

The second project which called for the use of different data structures was an analysis of the impact of earthquakes on lifelines. For this project, lifelines were defined to include public or private facilities that are important in maintaining public safety and economic activity. They include facilities such as electric power substations, which (from a regional perspective) can be defined as a point. Others (highways, railways, rapid transit lines, bridges and tunnels, power lines, petroleum pipelines) are most naturally represented as vectors.

The objective of this project was to overlay these lifelines on maps (data sets) containing projections of groundshaking intensity. The intensity data, which was derived from a series of models incorporating fault locations and geology, was stored in the hectare grid cell base. The lifelines were represented as points and vectors and then overlayed on the cell data to determine spatial coincidence. This procedure was used to determine both an average level of risk for each entity (such as a freeway segment) and to pinpoint specific parts of a network most vulnerable to damage. For some of the elements such as bridges, a ranking was established to guide priority for retrofitting.

A third illustration of new data structures is the use of GBF/DIME files in conjunction with other BASIS data. The DIME files can be seen as a street map in digital form (Figure 3). Each record is equivilent to a street segment; it contains street name and type, address range, identifiers for census and other geographic units on either side of the street, and coordinates for the end points. Transformation of these coordinates is used to put the DIME data in a spatial framework compatible with the other BASIS data. It can then be used to overlay street patterns on a map, or (with appropriate address-matching software) to estimate the location of a given address. These applications treat the DIME files as vectors or as a way to derive point locations.

This technique is limited by several practical considerations. File coverage is often incomplete and several years out of date, so some (often extensive) editing must be performed. Also, the spatial accuracy of the files must be carefully assessed, since questionable maps were often used for the original encoding. Despite these limitations, the DIME files are a unique data source with potential for many new applications.

FIGURE 3

SAN FRANCISCO STREETS

CONCLUSIONS

Our experience with BASIS has shown that it is indeed possible to incorporate multiple data structures in a regional data base. The three applications outlined above produced more useful results because of the ability to use different types of data in the most appropriate form for each.

There is some practical limit to the incorporation of multiple data structures in any one system, of course. The capabilities developed in BASIS do not allow for all possible (or perhaps even the best) data structures. There is a need to develop better techniques to derive the spatial interrelationships among all different modes. A common coordinate base is a necessary condition, but the next set of requirements (choosing the best mode of representation to use for any particular data set or application) needs more development.

The biggest benefit of multiple data structures is the ability to more effectively integrate data. In times of increasing demands and declining resources, this capability will become more and more necessary.

REFERENCES

The best sources of information about data structures for geographic data handling are the proceedings of several conferences in the field (or in related fields). Papers from previous Auto-Carto meetings are an excellent starting point. Other conferences include the annual Harvard Computer Graphics Week, ACSM/ASP meetings, and SIGGRAPH.

There are few books which cover this specific area. Mark S. Monmonier's Computer-Assisted Cartography: Principles and Prospects (Prentice-Hall, 1982) is one basic source. Standard references in the computer graphics field can be useful in providing comparisons of vector and raster structures.

Descriptions of BASIS applications are available from Geogroup Corporation or from the Association of Bay Area Governments. Several ABAG reports on earthquake mitigation are particularly good illustrations of GIS techniques being used with different data structures. Other examples of BASIS applications which utilize multiple data structures include models of solid waste disposal sites, airport noise analysis, and listings of environmental constraints on vacant industrial sites.

ADAPTIVE GRIDS FOR GEOMETRIC OPERATIONS *

Wm. Randolph Franklin
Electrical, Computer, and Systems Engineering Dept.
Rensselaer Polytechnic Institute
Troy, New York, USA, 12181.

ABSTRACT

A new data structure, the adaptive grid, is described. It makes a variety of geometric operations more efficient. For example, given a large set of small line segments, very few of which actually intersect, this will find the intersections in time proportional to the expected number of intersections. It will also determine which region of a planar graph contains a given point, in time proportional to the average number of edge segments that are cut by a scan line. These primitive operations can be used to perform higher level operations such as map overlay more efficiently.

The adaptive grid is a regular grid that is superimposed on the data. The size of the grid is determined from the statistics of the data. For example, in the edge segment intersection case, the grid size is set so that the expected number of edge intersections in each grid cell is some constant, such as five, independent of the total size and complexity of the data. A data item, such as an edge, is not cut into two pieces if it crosses a cell boundary. Instead, two records of the form (cell number, edge number) are written to a data structure.

The adaptive grid has been implemented and tested several times. For example, fifty thousand edges can be tested for intersections in about six minutes on a Prime 750 midi-computer.

INTRODUCTION

Spatial and topological relationships are integral to cartography, and an efficient use of computer data structures is essential in automated cartography. A new data structure, the adaptive grid, is presented here. It allows the efficient determination of coincidence relationships, such as "Which pairs of edges intersect?" and enclosure relationships like "Which region contains this point?". A good reference for computer data structures is (Aho, 1974). For cartographic data structures, see (Peucker, 1975). Some other useful algorithms from the area of computational geometry are (Bentley, 1979), (Bentley, 1980), and (Dobkin, 1979).

* This material is based upon work supported by the National Science Foundation under Grant No. ECS 80-21504.

DATA STRUCTURE

The adaptive grid data structure (in 2-D) is based on a single level uniform grid superimposed on the data. For example, suppose that we are given N small straight line segments or edges scattered within a square of side one. See figure 1, where N=4.

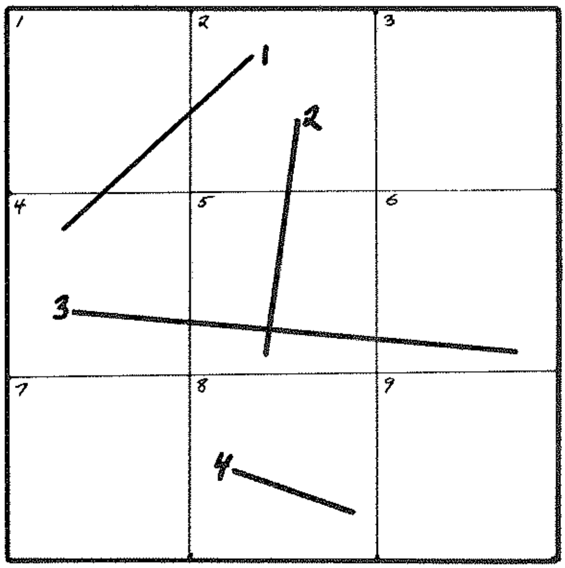


Figure 1: An Adaptive Grid Superimposed on 4 Edges

The grid has G by G cells (G=3 in figure 1), each of side B=1/G. Let L be a measure describing the edges' length, such as the average. The data structure consists of a set of ordered pairs of a cell number and an edge number, i.e.

{ (cell, edge) }

Each pair represents an edge passing through a cell. For figure 1, we obtain the following data structure:

{ (2,1), (1,1), (4,1), (2,2), (5,2), (4,3), (5,3), (6,3), (8,4) }

This data structure bears a superficial resemblence to several others: Unlike Warnock's hidden line algorithm (see Sutherland, 1974), the adaptive grid does not clip an edge into several pieces if it passes through several cells. In the example, edge #1 passes through cells #2, 1, and 4, but we merely obtain 3 ordered pairs in the set. Unlike a tree (see Nievergelt, 1974), or k-d tree (see Bentley, 1975a, and Willard), it divides the data cordinate space evenly independently of the order in which the data occurs. Unlike a quad tree (see Bentley, 1975b, and Finkel, 1974) or octree (see Meagher, 1982), the adaptive grid is one level, and does not subdivide in the regions where the data is denser. Although such a hierarchy is the most obvious "improvement" to an adaptive grid, it will be shown later that if the data is reasonably distributed, then a hierarchy would not increase the speed. Further, such a hierarchy would make it harder to determine which cells a given edge passes through.

When the adaptive grid is described as a set, the term is used precisely. The only operations to be performed on it are:

i) Insert a new element, and

ii) Retrieve all the elements in some order so that they may be sorted, adjacent elements combined under certain circumstances, and a new set created.

This gives a great freedom in implementing this abstract data structure. For example, on a microcomputer, the adaptive grid can be a sequential disk file, assuming that a file sorting routine exists.

FINDING INTERSECTIONS

We will now see how to determine all the pairs of edges that intersect. The operations are as follows:

i) Determine the optimal G, or resolution of the grid, from the statistics of the input edges. This will be described in more detail later, but letting G=1/L is reasonable. Initialize an empty grid data structure for this G.

ii) Make a single sequential pass through the input edges. For each edge, determine which grid cells it passes through, and for each such cell, insert a (cell, edge) pair into the grid data structure.

   Since determining exactly which cells an edge passes through requires an extension of the Bresenham algorithm (see Foley, 1982), which is a little complicated, in practice, an enclosing box is placed around the edge. The edge is considered to pass through all cells in this box. Considering an edge to be in some extra cells speeds up this section and slows down the pair by pair comparison in section (v) below.

iii) Retrieve the (cell, edge) pairs and sort them by cell number.

iv) Make a sequential pass through the sorted list. For each cell mentioned in the list, determine all the edges that pass through that cell, and combine them into a set. Now we have a new set:

$$\{ (\text{cell}, \{\text{edge}, \text{edge}, \ldots\}) \}$$

   with one element for each cell that has at least one edge passing through it. Each element has the cell number and a smaller set of the edges that pass through that cell. Continuing the example of figure 1, we have at this stage:

```
{ (1, {1}   ),
  (2, {1,2} ),
  (4, {1,3} ),
  (5, {2,3} ),
  (6, {3}   ),
  (8, {4}   ) }
```

Note that since cells 3, 7, and 9 have no edges passing through them, they do not appear at all here. An empty cell does not use even one word of storage.

v) For each element of this set, i.e. for each cell with at least one edge passing through it, test all the edges in the cell pair by pair to determine which intersect. Since two edges that intersect must do so in some cell, and so must appear together in that cell, this will find all intersections. In the example, edges 1 and 3 both pass through cell 4, but an exact test shows that they do not intersect. On the other hand, cell 5 contains edges 2 and 3, and they do intersect.

vi) If a pair of edges that do intersect appear together in more than one cell, then that intersection will be reported for each such cell. To avoid this duplication, when an intersection is found, it can be ignored unless it falls in the current cell. For this strategy to work, the cells must partition the space exactly, i.e. each point must fall in exactly one cell. This can be satisfied by considering each vertical grid line between two cells to be inside its right neighbor, and considering each horizontal grid line between two grid cells to be inside its upper neighbor.

TIMING

This method is useful only because it executes efficiently. We will now analyze the time and determine the optimal G.

Let U = the average number of cells that each edge falls in. Then, approximately

$$U = 1 + 2\,LG$$

However, assuming that we place a box around the edge and count all cells in that, as described above, we will get a higher figure:

$$U \le (1 + L/B)^2 = (1 + LG)^2, \qquad \text{since } BG = 1$$

We will use this higher number since it is more conservative

and does not affect the rate of growth of the final time.
Next, V = total number of (cell, edge) pairs

$$= N\ U$$

$$= N\ (\ 1 + LG\ )^2$$

Then, W = the average number of edges in each cell

$$= V\ /\ G^2$$

$$= N\ (\ B + L\ )^2$$

For the execution times, we will use the notation $\Theta(x)$, which means proportional to x, as x increases.

Let T1 = the time to calculate the (cell, edge) pairs

$$= \Theta(V)$$

and T2 = the time to determine the intersections

$$= \Theta(G^2\ W^2)$$

since in each cell, each edge must be tested against each other edge.
So T = total time

$$= T1 + T2$$

$$= \Theta(\ N\ (1 + LG)\ +\ N^2 G^2\ (B + L)^4\ )$$

Now, if we let $B=\Theta(L)$, i.e. the grid size is proportional to the edge length, we get

$$T = \Theta(N + S)$$

where

$$S = N^2 L^2\ /\ 4$$

since with Θ notation, which considers only the rate of growth, constant multipliers can be added freely. But S is approximately the expected number of edge intersections for a given N and L. Since a routine that finds all edge intersections must examine each edge and each intersection at least once, setting B=cL, for any c, gives an optimal time, up to a constant factor. The actual c which minimizes T should be determined heuristically, since it depends on the relative speeds of various parts of the program, and this may depend on the model of the computer.

We sometimes have even more freedom to select B, that is, there may be a range of functions for B that give the same minimum time, depending on the dependency of L on N as the problems get bigger. We can use this extra freedom to also minimize space if we wish.

It might be objected that the execution time for any cell depends on the average of the square of the number of edges in that cell, whereas we have used the square of the average. However, if the edges are independently and identically distributed, then each edge has the same independent probability of passing through any given cell. Thus the number of edges in a given cell is Poisson distributed, so the square of the mean equals the mean of the square.

IMPLEMENTATION

The adaptive grid has been implemented and tested in various applications. First, the edge intersection algorithm described above was implemented as a Flecs Fortran preprocessor (see Beyer, 1975) program on a Prime midi-computer. The largest example had 50,000 edges and 47,222 intersections. See table 1 for a list of execution times.

N	L	B	S	T1	T2	T
100	.100	.100	15	.17	.26	.43
300	.100	.100	153	.54	.93	1.47
1000	.010	.010	11	1.73	3.62	5.35
1000	.030	.030	163	1.72	2.54	4.25
1000	.100	.100	1720	1.71	4.46	6.18
3000	.010	.010	149	5.24	8.05	13.29
3000	.030	.030	1487	5.41	8.82	14.22
3000	.100	.100	15656	5.19	27.93	33.12
10000	.003	.010	156	16.36	16.45	32.82
10000	.010	.010	1813	17.38	26.02	43.40
10000	.030	.030	16633	17.68	44.78	67.45
30000	.001	.010	149	48.33	43.95	92.28
30000	.003	.010	1797	48.46	54.21	102.66
30000	.010	.010	16859	52.85	98.93	151.78
50000	.001	.010	315	77.71	75.75	153.46
50000	.003	.010	4953	79.49	92.37	171.87
50000	.010	.010	47222	86.23	278.49	364.72

<u>Column Labels</u>

N Number of edges
L Average length of edges, assuming screen is 1 by 1
B Length of side of each grid cell
S Number of intersections found
T1 CPU time (sec.) to put edges in cells
T2 CPU time to find the intersections among the edges
T Total CPU time

Table 1: Execution Times for Edge Intersections

From the table, we see that this program takes about (N+S)/300 CPU seconds to execute, even for the largest case. Some other facts about the largest case tested are: 98,753 (Cell, Edge) pairs, and 11,534 duplicate intersections.

Figure 2 shows the case with N=1000, G=10, L=0.1

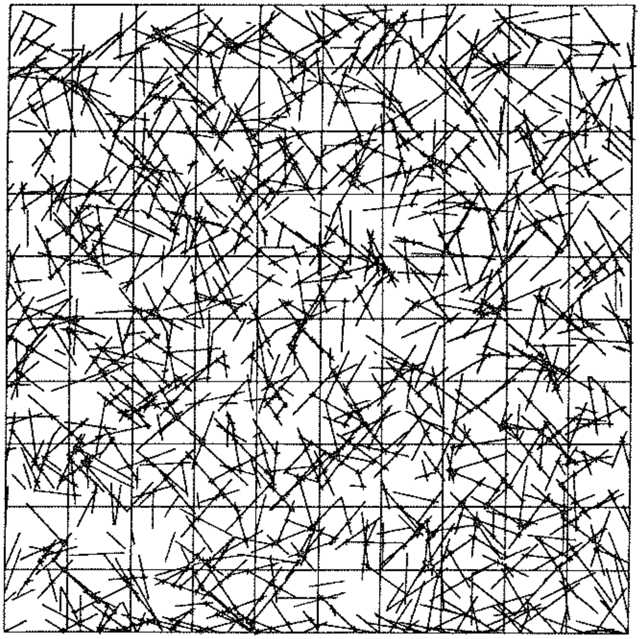


Figure 2: 1000 Intersecting Edges

approximately. In these examples, the coordinates of the edges are generated with a pseudo-random number generator. Now, manufacturer supplied random number generators are all linear congruential which has the effect that if you pair up successive random numbers, then the resulting 2-D points will fall on a comparatively small number of parallel lines. This does not mean that these edges will be parallel if a linear congruential generator is used since the random numbers are used for x1, y1, and the angle of inclination. Still, it is better to use a non-linear generator.

The adaptive grid was next used in a haloed line program designed and implemented by Varol Akman (see Akman, 1981, and Franklin, 1980). A haloed line drawing is a means of displaying a 3-D wire frame model (i.e. edges but not faces) of an object with front lines cutting gaps in read lines where they cross. This was tested on scenes with over 10,000 edges. It can process such a scene in about 10 minutes on a Prime, depending on the edges' lengths.

The Spheres program (see Franklin, 1981), is a third test. Here spheres are projected on top of one another. A case of ten thousand spheres of radius 0.02, overlaid on the average 10 deep through the scene, could be processed in 6.4 seconds. Here not only the intersections were determined, but also the visible segments of each sphere's perimeter were found.

Finally, the adaptive grid is an essential part of the simplified map overlay algorithm (see Franklin, 1983), currently under implementation.

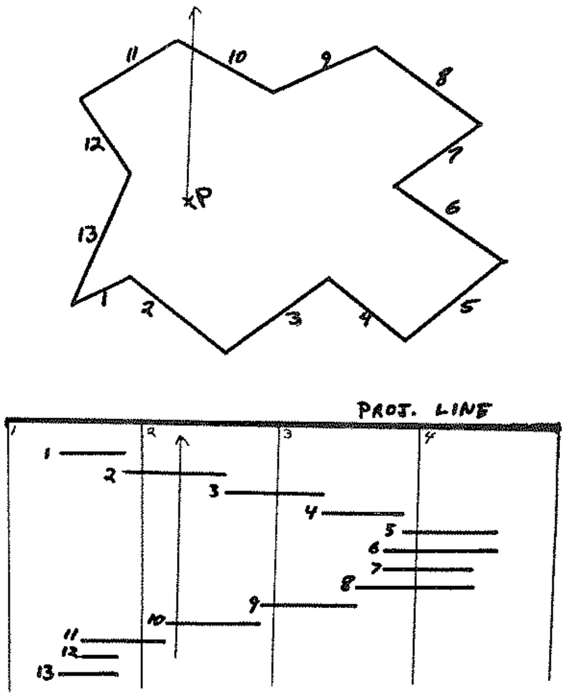


Figure 3: Point in Polygon Testing

TESTING WHETHER A POINT IS IN A POLYGON

The adaptive grid can also be used to test whether a polygon contains a point. If we preprocess the polygon first, this method is very efficient even if the polygon has many edges. The execution time per point depends on the depth complexity of the polygon, i.e. the average number of edges that a random scan line would cut. The total number of edges has no effect on the time.

This method is an extension of the method where a semi-infinite ray is extended from the point in some direction to infinity. The point is in the polygon if and only if the ray cuts an odd number of edges. The problem lies in testing the ray against every edge.

To speed this up, we will use a one dimensional version of the adaptive grid on a line. The line is divided into 1-D grid cells and then the polygon's edges are projected onto the line. Now we know which edges fall into each cell. For example, see figure 3, where a polygon with N=13 edges is projected onto a 1-D grid with G=4 cells. After all the edges in each cell are collected, we know, for example, that cell #2 has edges #2, 3, 9, and 10. Now, consider point P: Since it also projects into cell #2, a ray running vertically up from it can only intersect those 4 edges, so we need only test it against them. The execution time is the average number of edges per cell. As the cell size becomes smaller than the edge size, this number has the polygon's depth complexity as a lower bound.

LOCATING A POINT IN A PLANAR GRAPH

The obvious extension of the above problem is to take a planar graph and a point, P, and to determine which polygon of the graph contains P. This can be done by testing P in turn against each polygon in turn, unless one polygon completely contains another, but that is slow. A more efficient method is this:

i) Extend a ray up from P.

ii) Record all the edges that it crosses, along with those edges' neighboring polygons.

iii) Sort those edges by their ray crossings from P.

iv) Finally, P is contained in the lower polygonal neighbor of the closest crossing edge to P.

As before, we put the planar graph's edges into a 1-D adaptive grid and test the ray against only those edges in the same cell as P's projection.

SUMMARY

Techniques from computational geometry have been shown, both by theoretical analysis and by implementation, to lead to more efficient means of solving certain common operations in automated cartography.

REFERENCES

Aho, A.V., J.E. Hopcroft, and J.D. Ullman. 1974. The Design and Analysis of Computer Algorithms, Addison-Wesley, Reading, Mass.

Akman, V. 1981, HALO - A Computer Graphics Program for Efficiently Obtaining the Haloed Line Drawings of Computer Aided Design Models of Wire-Frame Objects, User's Manual and Program Logic Manual, Rensselaer Polytechnic Institute, Image Processing Lab.

Bentley, J.L. 1975a, "Multidimensional Binary Search Trees Used for Associative Searching", Comm. ACM 18(9), pp. 509-517.

Bentley, J.L. and D.F. Stanat. 1975b, "Analysis of Range Searches in Quad Trees", Information Processing Letters , 3(6), pp. 170-173.

Bentley, J.L. and T.A. Ottmann. 1979, "Algorithms for Reporting and Counting Geometric Intersections", IEEE Trans. on Computers, C-28(9), pp. 643-647.

Bentley, J.L. and D. Wood. 1980, "An Optimal Worst Case Algorithm for Reporting Intersections of Rectangles", IEEE Trans. on Computers, C-29(7), pp. 571-576.

Beyer, T. 1975, Flecs: User's Manual, Dept. of Computer Science, University of Oregon.

Dobkin, D. and R.J. Lipton. 1976, "Multidimensional Searching Problems", SIAM J. Comput., 5(2), pp. 181-186.

Finkel, R.A. and J.L. Bentley. 1974, "Quad Trees: A Data Structure for Retrieval on Composite Key", Acta Inform., 4, pp. 1-9.

Foley, J.D. and A. van Dam. 1982, Fundamentals of Interactive Computer Graphics, Addison-Wesley, Reading, Mass.

Franklin, W.R. 1980, Efficiently Computing the Haloed Line Effect for Hidden-Line Elimination, Rensselaer Polytechnic Institute, Image Processing Lab, IPL-81-004.

Franklin, W.R. 1981, "An Exact Hidden Sphere Algorithm That Operates in Linear Time", Computer Graphics and Image Processing, 15, pp. 364-379.

Franklin, W.R. 1983, A Simplified Map Overlay Algorithm, presented at Harvard Computer Graphics Conference, Cambridge, Mass., August 1983.

Meagher, D.J. 1982, The Octree Encoding Method for Efficient Solid Modelling, Ph.D. thesis, Rensselaer Polytechnic Institute.

Nievergelt, J. 1974, "Binary Search Trees and File Organization", ACM Computing Surveys, 6(3), pp. 195-207.

Peucker, T.K., and N. Chrisman. 1975, "Cartographic Data Structures", The American Cartographer, 2(1), pp. 55-69.

Sutherland, I.E., R.F. Sproull, and R.A. Schumacker. 1974, "A Characterization of Ten Hidden Surface Algorithms", Computing Surveys, 6(1), pp. 1-55.

Willard, D.E., Informative Abstract: New Data Stuctures for Orthogonal Queries, Harvard University.

TERRAIN APPROXIMATION BY TRIANGULAR FACETS

Albin Tarvydas
Marshall Macklin Monaghan Ltd.
Don Mills, Ontario, Canada M3B 2Y1

ABSTRACT

A set of subroutines has been developed for various digital terrain modelling applications. The modern concept of approximating a surface by triangular facets has been adopted as the basic construct in this development.

An efficient algorithm for computing the Delaunay triangulation in a two dimensional space is presented.

The triangulation algorithm operates in the complex domain and has been coded in FORTRAN 77. It does not contain any system-dependent functions, thus is easily portable to different computer systems supporting complex operators.

This fundamental algorithm incorporates various ideas proposed in recent research work and generates either a convex or a bounded triangular network over a random data set.

It operates in linear time over uniformly distributed data at an average rate of 50 points per CPU second as timed on our VAX 11/780 mini-computer.

In actual use, stereo digitized data sets containing 10000 - 20000 points were processed in about 5 minutes, while 100000 points required some 21 minutes of CPU.

The triangulation algorithm generates a simple yet effective data base which may be integrated in different digital terrain modelling applications.

Specifically, for each triangle in the data base, pointers to the three adjacent triangles are provided.

The pointers provide a direct means of traversing the data base in any desired fashion. Thus, vertical information for any horizontally defined line can be determined efficiently.

Similarly, watershed limits can be traced by comparing the flow vectors in the adjacent triangles.

The implementation of this data base in our organization for digital mapping, transportation and environmental engineering and in other digital terrain modelling applications is elaborated.

Contour generation techniques and implementation of splines in tension or Bézier curves for contour smoothing applications is discussed.

The production procedures, the associated computer costs, the relevant practical experiences and observations are also presented.

REDUCTION OF DIGITAL AEROGEOPHYSICAL DATA TO A LINEAR MODEL

M. T. Holroyd
Resource Geophysics and Geochemistry Division
Geological Survey of Canada (*)

ABSTRACT

Digital aerogeophysical surveys (aeromagnetic, radiometric and various types of electromagnetic) are characterised by their high rates of data acquisition and the consequent large volume of the data sets. A multi-channel gamma-ray spectrometer, for example, has an acquisition rate in excess of two kilobytes/second resulting in data sets in excess of 1000 megabytes for even a moderately sized survey.

The data content and structure differs between the various survey data types. In order to generalise the compilation system such that common processes may be applied to different data types by the same set of software modules, the software must be data structure independent.

The lengthy compilation processes required to derive a final mappable data set from the original raw data require mostly sequential access to large groups of data values, although some direct access processes are involved.

A linear model for aerogeophysical data structure is proposed which is well suited to the data access needs of the compilation systems. Furthermore, data independence, the "data dictionary", and the data manipulation language are derived from intrinsic properties of the model rather than being unrelated, post-factum additions as is the case with the hierarchical, network and relational data structure models.

---

* Now with Dataplotting Services Inc., Toronto.

MATHEMATICAL CARTOGRAPHY

Scale Preserving Smoothing of Islands and Lakes 252
John Oomen and R.L. Kashyap

An Adaptive Method for Numerically Modelling Large Numbers of Irregularly Spaced Data 261
Richard Schiro and Glen Williams

About Cartographic Contouring with Computers 271
Pinhas Yoeli

A Mathematical Evaluation of Simplification Algorithms 276
Robert McMaster

Visual Versus Computerized Seriation: The Implications for Automated Map Generalization 286
Jean-Claude Muller and John Honsaker

Automated Detection of Drainage Networks from Digital Elevation Models 297
David Mark

Shape Representation by Rectangles Preserving Their Fractality 308
Pietro Dell'Orco and Marco Ghiron

Fractal Enhancement for Thematic Display of Topogically Stored Data 318
Marc Armstrong and Lewis Hopkins

Measuring the Fractal Dimensions of Surfaces 328
Mark Shelberg, Nina Lam and Harold Moellering

An Algorithm for Variable-Width Feature Representation (Abstract) 338
Barbara Bonnette

SCALE PRESERVING SMOOTHING OF ISLANDS AND LAKES*

B. John Oommen and R. L. Kashyap
School of Computer Science
Carleton University
Ottawa: K1S 5B6
Canada

Department of Electrical Engg.
Purdue University
W. Lafayette, IN: 47907
U.S.A.

ABSTRACT

A smoothed version of a polygon ξ is defined as a polygon which approximates ξ according to a given criterion and which simultaneously has no more edges than ξ itself. In this paper, a scale preserving smoothing algorithm is presented. The input to the algorithm is a polygon ξ and the output is its smoothed version ξ_ε. ξ_ε, which contains all the scale information that ξ contains, is called the Linear Minimum Perimeter Polygon (LMPP) of ξ within a tolerance of ε. Using the quantity ε the degree to which ξ_ε approximates ξ can be controlled. From the LMPP a representation for a polygon approximating ξ can be procured, which is invariant to scale and translation changes. Examples involving the Great Lakes of North America have been included which demonstrate the applicability of the smoothing technique. Elsewhere we have shown how this technique can be used for thinning closed boundaries (KSO83) and in the processing stage of a pattern recognition system (KSO82).

INTRODUCTION

Over the past two decades considerable research has gone into the study of the automated processing of images. In this context polygons have played a major role especially since the outer boundary of an object without holes can be approximated as a polygon. The advantages of such representations can be found, for example, in (PAV77, Chapter VII). Consider a typical image processing environment in which the picture of an object to be recognized, silhouetted against a background is given by a two-dimensional pixel array. Using any boundary tracking algorithm a polygonal representation for the shape of the object can be obtained. It is not uncommon that such a representation involves a polygon having many hundreds of edges (MON70, PAV77, RAM72). Since the time required for processing the polygon is dependent on the number of edges it possesses, this polygon is usually approximated using a smoothing technique , such as the split and merge technique or the linear scan technique. These techniques and their variants have been well described in Pavlidis (PAV77, pp. 161-184).

Useful as these techniques are, none of these techniques preserves all the scale information contained in the unsmoothed boundary. To clarify this assertion, let ξ and τ be two unsmoothed polygons with τ being a scaled version of ξ , the scaling factor being $k > 0$. Let ξ^* and τ^* be the corresponding smoothed versions, the smoothing being performed using any of the algorithms known in the literature. Even though τ is a scaled version of ξ, none of the currently available techniques

*Partially supported by a grant from the Office of Naval Research, U.S.A., (Grant No. 82K0360).

can guarantee that ξ^* is a scaled version of τ^*.

In this paper we propose a smoothing scheme which can indeed guarantee the preservation of scale information. The input to the scheme is a polygon ξ and its output is ξ_ε, the smoothed version of ξ, referred to as Linear Minimum Perimeter Polygon (LMPP) of ξ within the tolerance ε. The quantity ε, $0 \le \varepsilon \le 1$, is termed as the tolerance factor. The value $\varepsilon=0$ yields ξ_ε identical to ξ and as ε increases ξ_ε approximates ξ more and more crudely. Within reasonable limits of ε, ξ_ε indeed preserves the scale information in ξ and yields ε as a single control parameter by which the smoothing can be controlled.

A natural consequence of this technique is a representation for a smoothed version of a shape, which is invariant to changes in scaling and the translation of the coordinate system in which the original shape is drawn.

This scale preserving smoothing technique will have immense value especially in the area of cartography. Maps of islands and lakes can be drawn to great precision provided a small scale factor is used. However, if a miniature map of an original has to be obtained, a fair amount of smoothing will have to be done. A pertinent question is one of knowing whether the miniature has all the relevant scale information. None of the smoothing techniques known in the literature can guarantee the preserving of this information. The smoothing technique presented here can be used to smooth the original map and the smoother version can be miniaturized with the full confidence of the map being to the appropriate scale.

To demonstrate the properties of this technique, cartographic examples have been included in the paper. These involve the Boundaries of the Great Lakes of North America obtained from televised pictures taken from the National Geographic collection (NGM53).

In the next section we shall describe the Minimum Perimeter Polygon (MPP) of a polygon. We then proceed to define the Linear Minimum Perimeter Polygon (LMPP) and demonstrate its properties. We conclude the paper with examples of the use of the LMPP in smoothing maps.

THE MINIMUM PERIMETER POLYGON

Let ξ be an polygon specified as an ordered sequence of points in the plane as below.

$$\xi = \{P_i/i=1,..N\} \qquad (1)$$

Let δ_i be a prespecified circular or polygonal constraint domain in the neighbourhood of P_i. Let ξ' be any polygon specified by the sequence

$$\xi' = \{R_i/i=1,...N;\ R_i \in \delta_i\} \qquad (2)$$

ξ' is any approximation of ξ in which the point P_i is perturbed to a new location R_i within the domain δ_i. The polygon ξ^* which satisfies (2) and which has the minimum perimeter is called the Minimum Perimeter Polygon (MPP) of ξ. Sklansky et al (SCH72, SKK76) and Montanari (MON70) have suggested algorithms for computing the MPP when the disjoint domains δ_i are polygonal or circular respectively. When the constraint domains are all disjoint circles with radius γ, the MPP ξ^* satisfies the following:

$$\xi^* = \text{Argument} \min_{\xi'} [f(\xi')] \quad (3)$$

where $\xi' = \{R_i / i=1,\ldots,N;\ \ ||P_i - R_i|| \leq \gamma\}$ and $f(\xi') = \sum_{i=1}^{N} ||R_i - R_{i+1}||$ with $R_{N+1} = R_1$. A vertex R_i of ξ^* is called an active vertex if it lies on the boundary of the constraint disk. In such a case

$$||P_i - R_i|| = \gamma.$$

Montanari has proved (MON70) that the MPP *, possesses the following properties:

(i) ξ^* can be completely and uniquely specified by the sequence of active vertices.

(ii) Let R_j and R_k be the first active vertices on either side of a vertex P_i of ξ. If the line joining R_j and R_k intersects the circular constraint disk around P_i, the vertex R_i will not be an active vertex. In such a case the edges of the MPP through R_i are colinear and can therefore be merged. This is depicted in Figure 1.

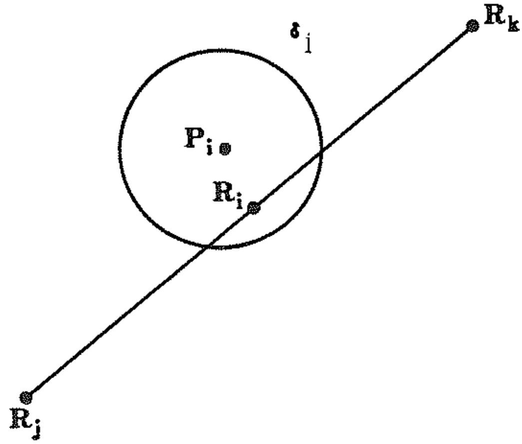


Figure 1.
The vertex R_i is not an active vertex of ξ^* since the line R_jR_k intersects the domain δ_i around P_i.

(iii) If the line joining R_j and R_k lies outside the constraint disk around P_i, the vertex R_i will be active. It will lie on the boundary of the disk at the point where the bisector of the angle $R_jR_iR_k$ is normal to the boundary. (See Figure II)

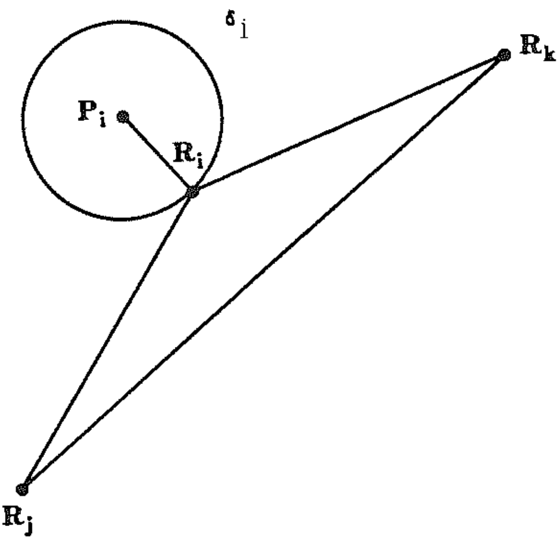


Figure II.
R_i is an active vertex of ξ^* since the line R_jR_k does not intersect the domain δ_i around P_i. The line P_iR_i bisects the angle $R_jR_iR_k$.

We shall now extend the results of Montanari to formulate the Linear Minimum Perimeter Polygon (LMPP).

THE LINEAR MINIMUM PERIMETER POLYGON

The Linear Minimum Perimeter Polygon (LMPP) of a polygon ξ is its Minimum Perimeter Polygon obtained by making the radii of the constraint disks directly proportional to the perimeter of ξ. The constant of proportionality ξ is called the tolerance factor, and the LMPP of ξ obtained using a tolerance factor of ε is given by ξ_ε. Specifically, if $\xi = \{P_i/i=1,\ldots,N\}$, ξ_ε is obtained by the following minimization procedure

$$\xi_\varepsilon = \text{Argument} \min_{\xi'} [f(\xi')] \tag{4}$$

with $\xi' = \{R_i/i=1,\ldots,N;\ ||R_i - P_i|| \leq \varepsilon L\}$

where $L = \sum_{i=1}^{N} ||P_i - P_{i+1}||$, $P_{N+1} = P_1$

and $f(\xi') = \sum_{i=1}^{N} ||R_i - R_{i+1}||$, $R_{N+1} = R_1$

Note that ξ_ε equals ξ if and only if ε is identically zero. We now demonstrate the scale preserving property of the LMPP. Let $|\mu|$ denote the perimeter of a polygon μ.

Theorem I
Let ξ be any closed boundary and ξ_ε be its LMPP within the tolerance ε. Let τ be a scaled version of ξ where the scaling factor is $k > 0$. Then τ_ε, the LMPP of τ within the same tolerance ε is the scaled version of

ξ_ε, scaled by the same scale factor k, provided the constraint disks are all disjoint.
The proof is included elsewhere (KSO83). ***

Remarks
(i) Since the uniqueness of the MPP requires that the constraint disks are disjoint ξ_ε preserves the scale information in ξ if and only if

$$\varepsilon < \underset{j}{\text{Min.}} [||P_j - P_{j+1}||]/2 \sum_{i=1}^{N} ||P_i - P_{i+1}||, \text{ where } P_{N+1}=P_1. \quad (5)$$

This range of ε is sufficient for some applications.

(ii) In many applications the above range of ε is not adequate. This is because of the fact that in many image processing applications two neighboring pixels may be adjacent vertices of a polygon which represents a boundary (FRE61, PAV77). In such cases, further smoothing can be achieved by preprocessing ξ to ensure that the constraint disks are disjoint. If P_k and P_{k+1} are two vertices in ξ obeying $||P_k - P_{k+1}|| \leq \varepsilon L$, a technique that we have found useful is to approximate ξ by merging P_k and P_{k+1} to a single point - for example, at their mean. The LMPP ξ_ε is computed using the approximated version of ξ. The latter approximation is pictorially shown is Figure III.

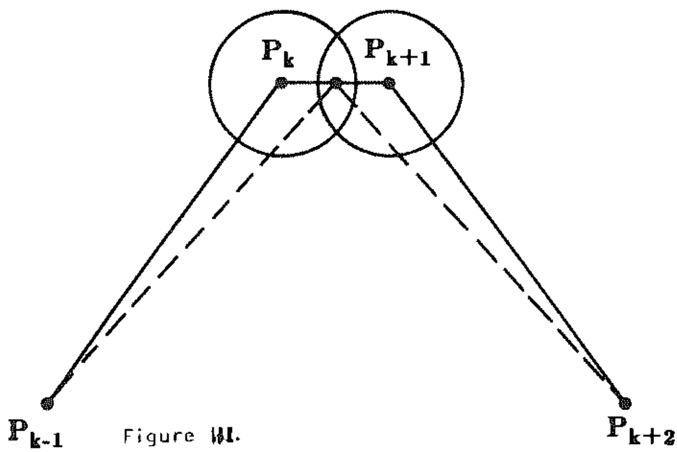


Figure III.

Approximating ξ to render the constraint disks disjoint. The bold lines are the original edges of ξ and the dotted edges result after P_k and P_{k+1} have been merged.

(iii) A consequence of Theorem I is that if ε satisfies (5), the LMPP provides a representation for a smoothed version of ξ which is invariant to scaling and translation of the original figure ξ. This representation is a sequence of angles and lengths, where the angles are those of the LMPP, and the lengths are the lengths of its edges normalized to its total perimeter. The use of this representation in pattern classification is discussed elsewhere (KSO82).

(iv) A note comparing the technique introduced here with the techniques currently used is not out of place. Various smoothing algorithms which use error norms have been proposed in the literature. Examples of these are the linear scan and the split and merge techniques (PAV77) and the method due to Ramer (RAM72). From a naive perspective it appears as if our contribution is merely that of introducing an error norm which is proportional to the perimeter of the polygon to be smoothed. Rather, we have exploited the properties of the active and nonactive vertices of the MPP and the uniqueness of the MPP. The error norm introduced here is augmented by the latter properties of the MPP to yield the scale presentation property of the LMPP claimed in Theorem I. In other words, we contend that if an error norm proportional to the perimeter is used in conjunction with other techniques such as the Linear scan or the split and merge technique, the scale information of the original polygon will be destroyed primarily because of the non-uniqueness of the resulting polygon.

(v) Ramer (RAM72) has noted that a smoothing technique should result in a polygon which possesses a minimal number of edges and simultaneously contains all the "significant features" of the original polygon. The LMPP possesses both these properties. Since all the edges which are "almost collinear" are merged the number of edges is drastically reduced. Further, since the vertices of LMPP fall on the boundaries of the constraint disks, the vertices which primarily distinguish the shape of the polygon are preserved, as in Figure II.

Example I

Let τ be the quadrilateral ABCD, given in Figure IV. The radii of the constraint disks around these vertices are given by L, where

$$L = || AB || + || BC || + || CD || + || DA ||.$$

The LMPP of τ is τ_ϵ given by the triangle PQR. Let the angles of the triangle be α_p, α_q, and α_r at the vertices P, Q and R respectively, and let the lengths of the edges opposite to P, Q and R be p, q and r respectively. Then, if Σ=p+q+r, the sequence of angles and lengths of the normalized LMPP will be the ordered sequence given below.

$$\{(\alpha_p, p/\Sigma), (\alpha_q, q/\Sigma), (\alpha_r, r/\Sigma)\}$$

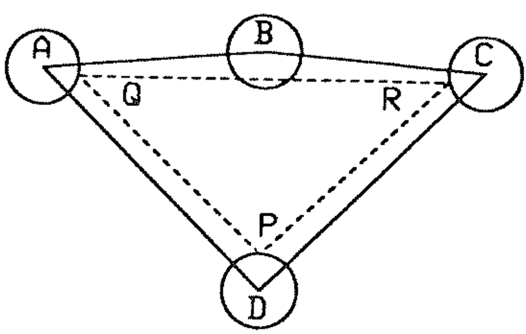


Figure IV.

The LMPP of the quadrilateral ABCD is the triangle PQR. The radii of the constraint disks are ϵ(||AB|| + ||BC|| + ||CD|| + ||DA||).

APPLICATIONS OF THE LMPP

A scale preserving smoothing technique will have immense value especially in the area of cartography. As observed earlier, maps of islands and lakes can be drawn to great precision provided a small scale factor is used. However, if the dimensions of the map are to be reduced, a fair amount of smoothing will have to be done. The researcher is now posed with the question of whether the miniature will contain all the relevant scale information. Using the LMPP to smooth the original map guarantees that if the smoothed version is reduced, the latter will indeed preserve the necessary scale information.

To demonstrate this feature we have considered the Great Lakes of North America. The maps were obtained from the National Geographic Magazine collection drawn at a scale of 32 miles/inch. They were appropriately scaled to fit an 8' by 11' frame. These pictures were then photographed using a television camera to fit a 90 x 90 pixel array. A simple boundary tracking algorithm utilizing constant thresholds was employed to extract the boundaries of the lakes.

The boundary of a map was used as the original polygon and the LMPP was constructed as its smoothed version. The effect of smoothing using various values of ε are shown for Lake Erie in Figure V (a) - (e). The original is shown in Figure V(a) and the smoothed versions for ε values of 0.001, 0.002, 0.005 and 0.01 are shown in Figure V (b) - (e) respectively. It can be seen that a value of ε=0.001 gives us a very fine approximation, and the approximation increases in crudeness as ε increases. A recommended value of ε for most applications is 0.005. This value of ε usually preserves the principal vertices of ξ, and reduces the number of edges by about 60 percent.

The applicability of the technique to perform thinning has been demonstrated by the authors (KSO83). Smoothing using the LMPP has been used in the preprocessing stage of a pattern recognition system. The details of this system are given elsewhere (KSO82).

CONCLUSIONS

In this paper we have presented a smoothing technique which has some interesting scale preserving properties. If ξ and τ are two polygons with τ as a scaled version of ξ, their smoothed versions ξ_ε and τ_ε are exactly scaled versions of each other and the scale factor is the same provided ε obeys a simple inequality constraint. The smoothed version known as the Linear Minimum Perimeter Polygon (LMPP) is a method of approximating ξ by ξ_ε where the latter has a number of edges less than or equal to the former. It also gives us a single parameter ε to control the degree of approximation.

A consequence of this is that we can represent a polygon ξ approximately by a string of real number pairs and this string is invariant to the scale and the coordinate system of ξ.

Experimental results which demonstrate the smoothing performed by the LMPP have also been presented. These results involve the boundaries of the Great Lakes of North America obtained from the National Geographic Magazine (NGM1953).

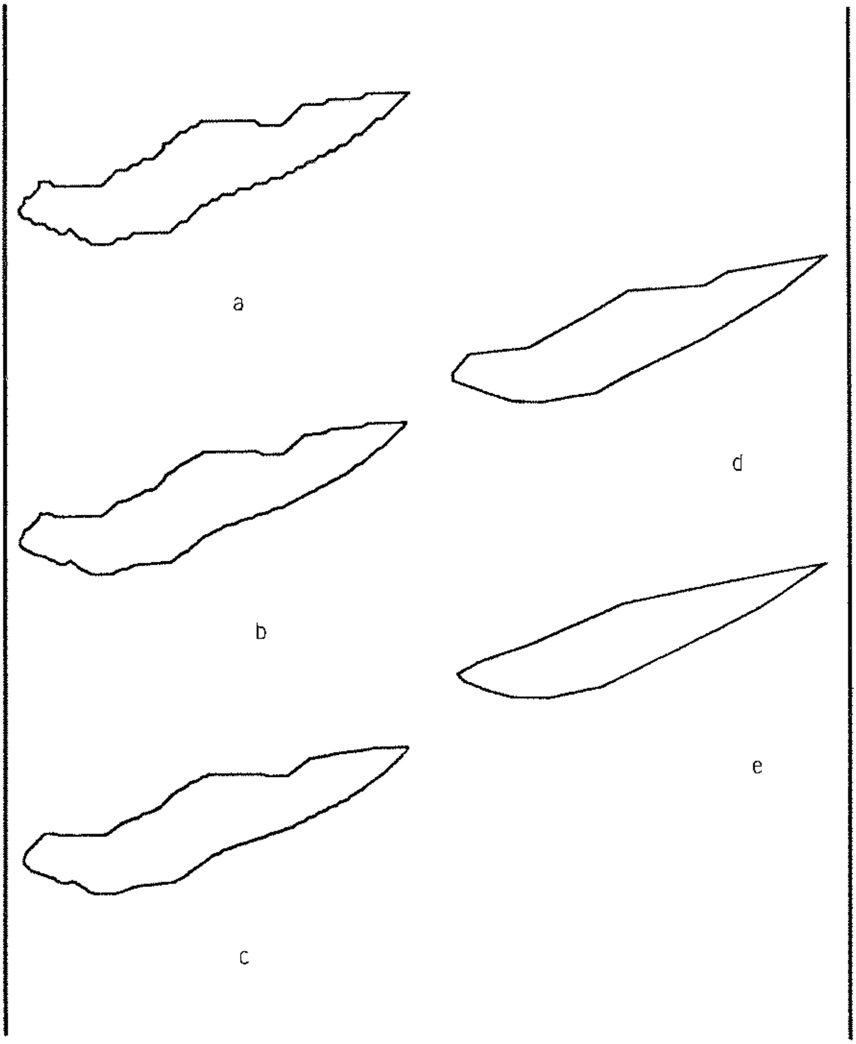


Figure V .

Examples of smoothing the boundary of Lake Erie. (a) is the original and (b)-(e) are the LMPPs drawn using ε values of 0.001, 0.002, 0.005 and 0.01 respectively.

REFERENCES

FRE61 Freeman, H. 1961, "On the Encoding of Arbitrary Geometric Configurations", IRE Trans. on Computers, Vol. EC-10, pp. 260-268.

KSO82 Kashyap, R. L., and Oommen, B. J. 1982, "A Geometrical Approach to Polygonal Dissimilarity and the Classification of Closed Boundaries", IEEE Trans. on Pat. Anal. and Machine Intel., Vol. PAMI-4, pp. 649-654.

KSO83 Kashyap, R. L., and Oommen, B. J. 1983, "A Scale Preserving Smoothing Technique", IEEE Trans. on Pat. Anal. and Machine Intel., Vol. PAMI-5. (Issue not known).

MON70 Montanari, U. 1970, "A Note on Minimal Length Polygonal Approximation to a Digitized Contour", Comm. of the ACM, Vol. 13, pp. 41-47.

NGM53 The National Geographic Magazine 1953, Maps of the Great Lakes Region. Drawn Dec. 1953.

PAV77 Pavlidis, T. 1977, Structural Pattern Recognition", New York: Springer-Verlag, 1977.

SCH72 Sklansky, J., Chazin, R. L. and Hansen, B. J. 1972, "Minimum-Perimeter Polygons of Digitized Silhouettes", IEEE Trans. on Comp., C-21, pp. 260-268.

SKK76 Sklansky, J., and Kibler, D. F. 1976, "A Theory of Nonuniformly Digitized Binary Pictures", IEEE Trans. on Systems Man and Cybern., SMC-6, pp. 637-647.

RAM72 Ramer, A. 1972, "An Iterative Procedure for the Polygonal Approximation of Plane Curves", Computer Graphics and Image Processing, pp. 244-256.

AN ADAPTIVE METHOD FOR NUMERICALLY MODELING LARGE NUMBERS OF IRREGULARLY SPACED DATA

LCDR Richard A. Schiro
National Oceanic and Atmospheric Administration
Pacific Marine Center, MOP2X3
1801 Fairview Avenue East
Seattle, Washington 98102
and
Glen Williams
Data Processing Center
Texas A & M University
College Station, TX 77843

This paper explores one technique for applying multiquadric interpolants to large numbers of irregularly spaced two dimensional data, e.g. hydrographic data. The technique gives a model which locally approximates in an ℓ^∞ sense or locally interpolates the original data. The technique was applied to five representative sets of hydrographic data, consisting of over 12,000 points, and two sets of data computed from common, mathematically defined surfaces in the literature. For the largest set of data which was comprised of over 3000 points, the partitioning and grouping required 28 seconds, the modeling required 11.5 minutes, and the computation of a regular grid from the model consisting of 20,000 points required 3 minutes of AMDAHL 470 V/6 CPU time.

INTRODUCTION

Since the United States government can be held legally responsible for its hydrographic surveys as well as the individual soundings, an accurate model of a hydrographic survey could serve many practical and useful needs in nautical chart production. The three areas in which an accurate model would be useful are survey data reduction, survey verification, and the generation of accurate depth contours for nautical charts.

Table 1. Hydrographic Vertical Accuracy Requirements

Depth (Meters)	Accuracy (Meters)
0-20	0.3
20-100	1.0
100	1% of depth

At the present time the National Ocean Service's Automated Information System is becoming operational. It is estimated that of the three to four billion bytes of on-line data, 60-70% of that total is due to unpublished hydrographic soundings, which are soundings eligible to be charted but are not chosen for cartographic reasons. If a model could be constructed which would use only a subset of the original soundings and accurately predict the remainder, then a significantly reduced amount of data could be stored, resulting in considerable human/computer resource savings.

Currently as a new survey is being conducted, comparisons are made with all previous charts, older surveys, and junctioning contemporary surveys. This is necessary to ensure accurate data are being collected, and to identify areas needing more development due to previously undetected dangers to navigation. Since adequate numerical computer models do not exist for this type of data, these comparisons are currently made by hand. This can lead to errors both in accepting erroneous data as well as missing opportunities to investigate newly detected dangers. If an accurate model was available for a prior survey or chart, then the verification of hydrographic surveys could be performed in a more quantitative and consistent manner.

The generation of depth contours requires a surface to be generated or grid points to be estimated from the original discrete data. On a nautical chart the depth contours as well as some of the original discrete soundings are portrayed. If the surface or grid values from which the contours are derived do not closely approximate or interpolate the original discrete data, the discrete data and the contours on the chart will have significant discrepancies. If an accurate model was available, then the surface or the estimated grid values could be used to generate more accurate contours.

PRESENT STATE OF MODELING IRREGULARLY SPACED DATA

The determination of a surface described by irregularly spaced data can be stated as:

Given the points (X_k, Y_k, Z_k), k=1,...,N over some domain (D(X,Y)), a function is desired which reproduces the given points and produces a reasonable estimate of the surface (Z) to all other points (X, Y) in the given domain.

There are two general approaches to surface modeling via irregularly spaced data - approximation and interpolation (Schumacher, 1976). Approximation is generally used when the data are considered to be "noisy" or have some measurement errors. Least square-based algorithms are common in these techniques. Interpolation is generally used when the data represent "exact" values and/or there is a need to ensure that the model reproduces the initial data.

Further, approximation or interpolation methods can be classified as being global or local. Global methods are those in which all of the data participate in determining the modeling function; local methods are those in which the modeling function is determined by data "nearby" that point where the function will be evaluated. Numerous interpolation methods exist for both global and local methods. Extensive bibliographies appear in (Franke, 1979) and (Schumacker, 1976).

Usually, local methods are applied to large numbers of data rather than global methods. This is normally due to the ill-conditioning mentioned above. However, global methods will be considered for hydrographic data for three reasons. First, the application of a global method can approach a local method as the numbers of data to which it applies decrease. Second, for large numbers of data, local methods require efficient access of large complex data structures. Third, a global method when employed as an ℓ_∞ approximation (constraining the maximum difference between the model and the original data) may lead to reduction of the data needed for modeling. This could be significant given large numbers of data.

(Franke, 1979) conducted numerous tests on twenty-nine methods including both local and global methods. He judged that the multiquadric interpolation method was the best global method tested. Specifically, he judged it excellent in terms of accuracy and visual aspects and moderately good in terms of computation time. Hein, 1979 conducted tests on eight interpolation methods and also concluded that the multiquadric technique was the best method tested.

THE MODELING TECHNIQUE

The Modeling Data Structures

Since the goal of this paper was to numerically model large numbers of irregularly spaced data, a standard algorithmic design technique was employed - divide and conquer. In place of one model for all of the data, there would be one model for each division of the original set of data. This required the design of a data structure requiring at least two levels. The first level must address how the whole has been divided, and the next levels of the data structure must address in increasing detail each of the divisions. For this technique two levels were designed. The first level, called a directory, describes the needed information on each division of the original set of data. It also contains information on how each division relates to the other divisions and information about the data in each division. The second level is divided in two parts - the model of each division and the original data assigned to each division. Each model of a division is a numerical model such that given an (X, Y) coordinate in a division, the depth, Z coordinate, can be computed. The original data in a division is all the data assigned to a division from which the model was computed.

The Directory. The directory describes the initial and final divisions of the domain of the original set of irregularly spaced data. These initial divisions are called cells. The user specifies the rectangular limits of the domain and the approximate number of cells desired in the model. From this the modeling software computes the numbers of rows and columns of cells depending on the rectangular limits of the domain, such that the domain falls entirely within the array of cells.

The cell directory is the key data structure and serves several important functions, a few of which are:

1. Given an (X, Y) pair, the cell to which it belongs can be easily computed via simple arithmetic.
2. From the cell directory an index table can be created to index directly to the appropriate multiquadric model.
3. It contains governing geometric and cell group parameters. Access to the cell directory is random by cell number, and software is provided for its storage in in-core tables or random access files.

Data Structures. The data structure for the individual models is purposely simplistic. The model output is a sequential file in which each model is output after its computation. If the method of computing with the model dictated access to a random model, or if it were known that the evaluations would be sequential as regards to model number, efficient access might require different structures. Also, the storage size of all the models taken together would influence the data structure choice. Thus the models could be imbedded in the application program or stored on random access storage. The model data structure used in this paper could be directly used to support either one. The data structure for the original data is also simple, and consists of a sorted sequential file with one point per record.

The Multiquadric Modeling Equations

A multiquadric surface can be written as

$$Z=\sum_{i=1}^{n} C_i[Q(X, Y, X_i, Y_i)] \tag{1}$$

in which each Q is a quadric surface, (Hardy, 1981) Several different quadrics can be used, a few of which are:

1. For a hyperbolic quadric

$$Q = [(X-X_i)^2 + (Y-Y_i)^2 + d^2]^{1/2} \tag{2}$$

2. For a recirpocal hyperbolic quadric

$$Q = [(X-X_i)^2 + (Y-Y_i)^2 + d^2]^{-1/2} \tag{3}$$

3. For a conic quadric

$$Q = [(X-X_i)^2 + (Y-Y_i)^2]^{1/2} \tag{4}$$

4. For a parabolic quadric

$$Q = [(X-X_i)^2 +/- (Y-Y_i)^2] \tag{5}$$

(Pickrell, 1979) ran rests on the hyperbolic, reciprocal hyperbolic, and conic multiquadric surfaces on hydrographic data, in which the conic quadric performed the best. Also, use of quadrics other than conic or parabolic types requires the computation of the parameter d. (Franke, 1979) noted that d = .815r worked best, where r is the average distance to the nearest neighbor. However, (Pickrell, 1979) found that if d is not zero then the surface is scale-dependent. For hydrographic and other types of data, this is not acceptable. For these reasons the conic quadric is recommended when scale dependancy is considered important, and in this paper the majority of computation was done with conics.

Additionally, the equation for Z in each model was modified to ensure that constant data would result in a constant as the only model term. Thus, the equation was modified to

$$Z = \sum_{i=1}^{n} C_i \left[(X-X_i)^2 + (Y-Y_i)^2 \right]^{\frac{1}{2}} + Zmean \qquad (6)$$

This will give the ability to reproduce constants and still preserve the symmetry property of the resulting matrix, thereby allowing the employment of an algorithm for the solution of the system of equations which exploits the symmetry and is efficient with regard to storage space and execution time.

Cell Directory Generation and Model Input Preparation

Given the rectangular limits of the domain to be modeled and the approximate number of cells desired, the algorithm will determine the initial numbers of rows and columns of cells depending on the shape of the domain to be modeled and the cell size necessary to totally cover the domain. At this point the cell directory is initialized with each cell's depth and standard deviation of depth, as computed from the input survey data.

An attempt is then made to form rectangular groups of one or more cells which are similar. These rectangular groups of similar cells are called kernel groups. Similarity in this application infers that all cells in the kernel group have mean depths which are within one standard deviation of the depth of the cell in the lower left hand corner of the kernel group. This cell is called the kernel cell.

The reason for attempting the grouping process is to attempt a greater data reduction capability. It was felt that if several areas are similar in some appropriate parameter, then it is more likely that one model could suffice for the entire kernel group.

Model Computation

The model computation uses an iterative approach on a kernel group-by-kernel group basis. For a specific kernel group, the algorithm initially selects model points and solves for the coefficients of the resulting multiquadric equation. All points of the kernel group not in the model are compared to the predicted values from the multiquadric model. On each iteration, some of the points out of tolerance are added to the model, and on every other iteration, the least critical points are removed for at least one iteration. This general approach was suggested by a manual selection/de-selection process by (Pickrell, 1979). In his work the iteration continued until no further significant improvement in the mean square error resulted. The overall procedure is an automated process and continues until all points are in tolerance.

Point Selection. The initial set of points are selected by computing the differences from the mean depth and selecting up to a specified number of the largest differences out of model tolerance. If all points are in model tolerance, then no additional points are selected and the model consists solely of the mean. Otherwise, the first point selected is the largest difference from the kernel group's mean. As each additional point out of model tolerence is considered in decreasing difference order, a test is made to determine if it is at least a certain distance from all previous points selected. If this distance criterion is satisfied, the point is selected. The above procedure is repeated until a desired number of points is selected, or until all points are considered.

RESULTS AND CONCLUSIONS

Description of Test Areas

Seven test areas were chosen for the iterative application of the multiquadric method in this paper. The first five of the test areas came from hydrographic survey H-8703, Swan Point to Dahlgren, Potomac River, Maryland and Virginia, and the areas were chosen to be representative of the majority of hydrographic surveys. The last two areas are surfaces defined by mathematical functions and referenced in the literature to compare approximation or interpolation methods. (See figures 1,2 and 3).

There were three reasons for choosing both real and generated data. First, it was desired to test the multiquadric methods on real sets of data which represent a variety of surface conditions and consist of large numbers of data. Second, the mathematical surfaces were used to compare the iterative multiquadric method using surfaces which were referenced in the literature. Third, the mathematical surfaces were used to test the model's behavior at points other than the original input data. It was felt that this was important since with real data, although the surface was and can be compared to the original data, the value of the surface at other points is unknown and often unknowable. Hence, by generating and then modeling data from a known surface, the model's behavior at points other than the original data could be tested after the fact.

Test Results

Extensive parametric testing of a software package which implements this modeling approach revealed the following solution criteria:

1. Conic quadrics outperform hyperbolic, reciprocal hyperbolic and parabolic quadrics for hydrographic data. This supports (Pickrell, 1979)'s conclusion.
2. To achieve an acceptable balance between data reduction and model computation time, the point selection parameters are defined as:
   a. Points to be added: 5% of the points in a kernel.
   b. Points to be deleted: 2.5% of the points in a kernel.

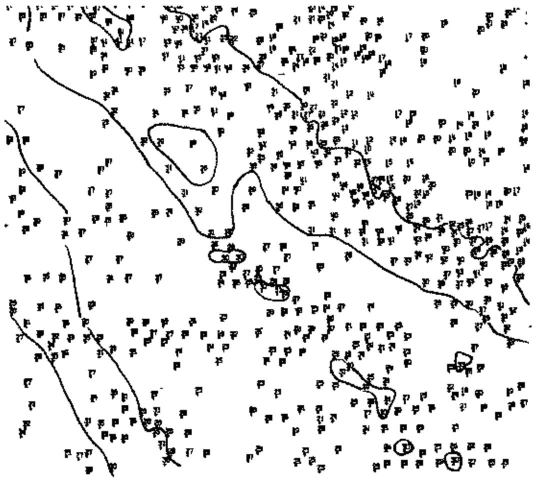


Figure 1. Point-Reduced Sample of Test Area 5

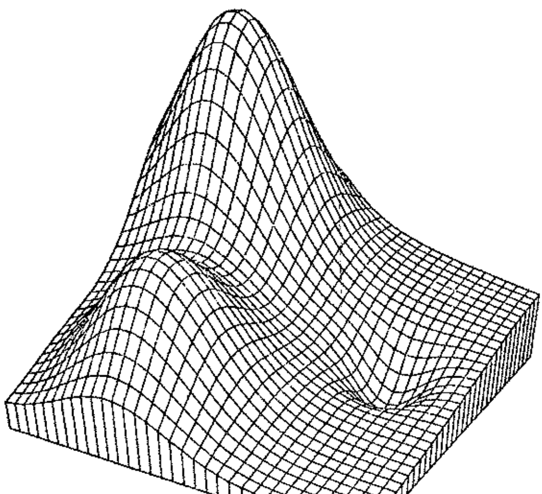


Figure 2. **Test Area 6** ~~**(Reproduced from Franke**~~ **79,** Figure 4.0.1.0)

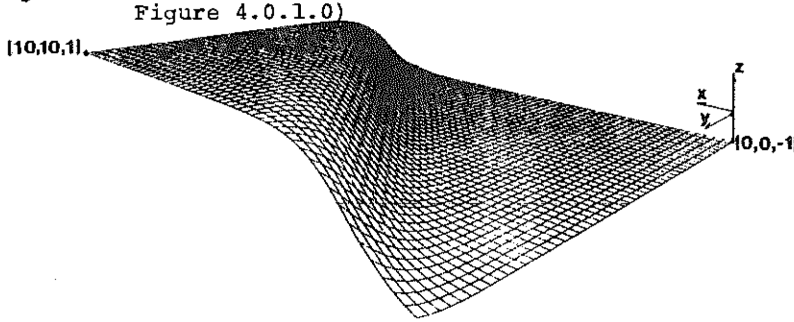


Figure 3. Test Area 7

c. Point addition distance criterion: Square root of the kernel area divided by the points in the kernel group.

Model Generation. All models were generated using conic multiquadrics. Areas 1 through 5 were run with the hydrographic error criteria (see Table 1), and Areas 6 and 7 were run with several absolute error criteria.

The results in Table 2 indicate that the method is feasible to model large numbers of irregularly spaced data. The major factors in determining the total running time are the solution techniques for solving the systems of equations, the number of iterations, and the variance of the data to be modeled.

The technique employed in solving the systems of equations is due to (Aasen, 1971). It is a stable algorithm for reducing a symmetric non-definite matrix of order N to tri-diagonal form, involving about $N^3/6$ additions and multiplications.

The number of iterations in the tests ranged from seven to twenty-three with the average around eleven. It is noted that the run parameters significantly affect the number of iterations. If data reduction is of primary concern, then a higher percentage of deletion points would be desired, and conversely for computation time.

Table 2. Results of Initial Cell Generation/Model Computation

Area	Number of Kernels	Initial Cell Generation (CPU Seconds)*	Model Computation (CPU Seconds)*
1	11	28	1130
2	5	23	494
3	2	12	62
4	9	26	1025
5	8	28	1400

* AMDAHL 470 V/6

If a very rough area is to be modeled, it is recommended that the grouping process not be attempted. On further tests with Area 5, it was determined that comparable overall data reduction was accomplished using only half of the computation time without combining cells as when pre-grouping was performed. This is due to cells being combined as similar when in fact they are similar only in their irregularity, hence time was spent trying to model the dissimilar areas, de-grouping cells and recomputing the models.

Data Reduction Properties of the Model. The most novel feature of this study is the approach to the error criteria as applied to the modeling technique. As was mentioned previously, all original data must be in the model or be predicted by the model to within a specified tolerance. For some types of data a least squares approach is not wholly applicable. Hydrographic data is one of these types, hence, this approach, a local ℓ_∞ norm approach, is more appropriate. Table 3 summarizes the results of the test areas.

Although the faithfulness of the model to the original data is guaranteed, the faithfulness of the model at other points had to be tested. For hydrographic data, the model for Area 4 was used to generate a contour map. The contours were very pleasant and compared favorably to hand-contoured plots and reasonably well to machine contouring with smoothing.

For the mathematical surface comprising Area 6 one hundred pseudo-random points were generated and a multiquadric surface was fit such that $|Z(X, Y) - \text{Model}| \leq 0.012$. This required 51 of the original 100 points. The surface model was then evaluated at an interval of 0.025 over the region [0,1]x[0,1]. The maximum difference was 0.09 and occured at the global maximum which was not represented in the original data.

It should be noted that the variance of the data has a major influence on the total computational time needed to compute the model. Areas 2 and 3 were the simplest followed by Areas 1, 4 and 5. Although this is not a startling observation, what is important to note is that the total computational time is not directly related to the total number of points, N, but related to the total number of reduced points in the model, M. Test results indicate that M is approximately N/2. The divide and conquer approach, coupled with the fact that not all N points are needed for modeling, can account for well over one or two orders of magnitude improvement in the time needed to solve the N-by-N system of equations.

Table 3. Results of Data Reduction for Test Areas

Area	Error Criterion Z(X, Y) - Model	Model Points	Total Points	Percent Reduction
1	Hydrographic	1378	2844	52
2	"	674	2231	70
3	"	216	1257	83
3	0.5 + .01* Z	391	1257	69
4	Hydrographic	1181	1641	55
5	"	1553	3027	49
6	0.2	10	100	90
6	0.07	19	100	81
6	0.012	51	100	49
7	0.1	50	300	84
7	0.01	145	300	52

CONCLUSIONS

The paper presents simple and relatively efficient data structures, techniques, and mathematical algorithms to employ a local ℓ_∞ approximation or local interpolation to large numbers of irregularly spaced data. Although approximations have been employed before in data fitting, the iterative selection technique, the enforcement of a specified error tolerance for all data and the attempted data reduction is a novel approach.

Additionally, this method performed well when tested with a large set of real data and some common mathematical test surfaces. The multiquadric model acceptably approximated the original mathematical surfaces even in areas which original data was not present. This approximation was done using only 50% of the original data. Hence, the tests show that the method can produce good results and be reasonably efficient even with rather strict error criteria.

REFERENCES

Aasen, J.O. 1971, On the Reduction of a Symmetric Matrix to Triadiagonal Form: Nordist Tidskrift For Information - sbehandlung, Vol. 11, pp. 233-242

Franke, R. 1979, A Critical Comparison of Some Methods for Interpolation of Scattered Data: Report Number NPS-53-79-002, Naval Postgraduate School, Monterey, California

Hardy, R.L. 1981, Comparative Studies of Multiquadric Equations and Other Methods of Interpolation and Prediction: Technical Papers of the American Congress on Surveying and Mapping, 41st Annual Meeting, February 1981

Hein, G. and Lenze, K. 1979, "Zur genauigkeit und wirtshaftlichkeit verschiedener interpolations und pardiktionsmethoden", Zietschrift fuer Vermessungswensen, Vol 104, November, 1979, pp. 492-505

Pickrell, A.J. 1979, Representation of Hydrographic Surveys and Ocean Bottom Topography by Analytical Models: M.S. Thesis, Naval Postgraduate School, Monterey, California

Schumaker, L.L. 1976, Fitting Surfaces to Scattered Data: Approximation Theory II, Academic Press, New York

ABOUT CARTOGRAPHIC CONTOURING WITH COMPUTERS

P. Yoeli
Tel-Aviv University - Department of Geography
Tel-Aviv 69978 - Israel

ABSTRACT

Digital models of height points may serve to define an approximation surface to the true earth's surface or to depict the spatial distribution of other geographical or geological phenomena. A basic distinction between digital models relating to the earth's surface and those depicting thematic phenomena is made. The view is expressed that algorithms for the computer assisted contouring based on digital models should be adapted to the type and character of DTM's at hand and the advantage of the use of "general purpose" contouring algorithms is doubted. Some methodological remarks about the computer assisted generalisation of topographical relief are made.

Contours in the cartographic sense are the projection onto the map plane of lines connecting points of equal elevation on a three dimensional surface. They serve as a projected three dimensional graphic measuring device enabling the map reader to gather altimetric information from the map and to recreate the depicted surface either mentally or physically by elevating every contour to its noted height.

In the case of a mathematically defined surface determined by a formula which expresses the elevation z of any point on the surface as a function of its x and y coordinates ($z = f(x, y)$), the contours are the projection of lines connecting points of constant value of z. These are so-called "parameter lines" of the function. For the graphic depiction of such a surface the mathematical formula is enough to compute and draw any required contour system, either manually or computer assisted.

For the depiction of surfaces void of any mathematical regularity, such as the earth's relief, the only way to produce contours is either to survey them directly on the surface - if this is possible - followed by their graphic presentation, or to survey the spatial coordinates of a certain amount of points spread over the surface, to connect these points with a spatial grid and to relate to this three dimensional "construction" as an approximation surface to the real surface. The contours are then interpolated in between the points of this grid. The first option is often applied in tacheometric ground surveys where the course of the contours is found directly in the field, choosing on the ground only those points having the precise elevation of the contours of the desired contour system. A similar procedure, in principle, is applied in topographic contouring from the stereoscopic models of airphotos. With conventional topographical stereo-photogrammetry, a measuring mark appearing to be floating in space is moved along the slopes of the virtual stereoscopic model at the height of a contour. The transmission of this movement and its graphic depiction on a drawing medium then results in the desired contour.

For thematic maps in the various fields of the earth sciences and geography, contours were, and still are, almost always the result of interpolations into arrays of a discrete number of elevation points, while on topographical maps and especially on those of photogrammetric origin, contours were mostly the result of direct contour measurements. With the advent of computer-assistance in map making, and consequently in the contouring process, the analogous survey of contours of topographical relief has been replaced by the digital survey and registration of height points. Computer assisted relief cartography of both thematic and topographical maps is therefore based today on the existance of data files comprised of the coordinates of a selection of points of the surface to be contoured. These data files are commonly known as "Digital Terrain Models" - DTM's in short - but as the data for contours of thematic maps does not necessarily refer to "terrain", we propose to restrict the term "DTM" to the height points on the earth's surface, and use the name "Digital Elevation Models" (DEM) for all the other geographical and geological phenomena.

The geometrical disposition of the points defining the approximation surface to be contoured can be, in principle, on one of the following forms:

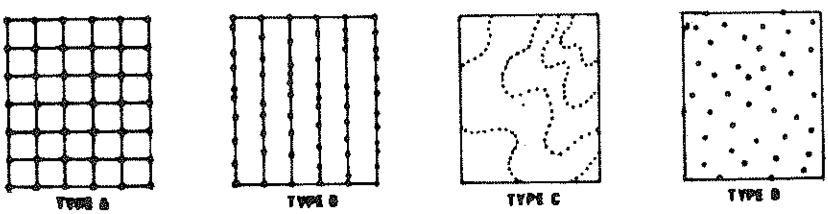


Fig. 1

A: Regular grid.
B: Points at irregular distances along equidistant parallel profile lines.
C: Arrays of points at equal heights.
D: Random point distribution.

The full diversity of types of point distributions occurs in the terrestrial and photogrammetric survey of the earth's surface, i.e., with "DTM's" and is the result of the various surveying methods applicable. Terrestrial surveys usually result in DTM's of type C or D while A and B are typical of photogrammetric surveys. Most "DEM's" serving as input for thematic contour maps are usually of type D.

It is most important to realise that the distinction between the various digital point models goes far beyond their different geometrical point distributions. There is a basic difference between the character of digital models of the earth's surface and those relating to thematic phenomena such as the distribution of temperature, or the results of bore holes of gravimetric surveys for the composition of geological or geophysical maps.

The aim of contours on a topographical map is to document the earth's surface as truthfully as the map scale allows. These contours are supposed to depict a reality not an assumption. To achieve this, the degree of the approximation of the surface defined by the DTM must be such, that the accuracy of the interpolated contours is in accordance with the tolerances stipulated for the specific contouring job at hand. This degree of approximation of the substitute surface is determined by the surveying process and depends on:

A. The accuracy of the point measurements;
B. The point density;
C. The point selection.

As the earth's surface is a real surface, visible, and in most cases also accessible, the quality of a surveyed DTM-terrestrial or photogrammetric - depends on the surveyor or the photogrammetric operator, and on them alone, and can be regulated by them at will. When using digital terrain models as input for the computer assisted generation of contours it must be assumed that the DTM was produced according to the requirements of professional surveying standards and that the originally surveyed points, and consequently the surface defined by them are the best available input. There is little sense in trying to "improve" an original DTM by computing what amounts to a secondary DTM comprised of computed points at a different geometrical disposition, instead of the original points. This is, however, done often in order to transform DTM's of type B, C or D into those of type A. The main reason for this, which in our opinion is a questionable procedure, is the wish to present the contour-threading computer program with an input of points arranged in a fully regular grid. It is, indeed, much easier to formulate a computer algorithm for the threading of contours into a regular array of points of type A than for any of the other possible point dispositions. (This includes type C where the logical problem is not, as one may think the connection of the given points with curves but rather the interpolation of a different contour system). To forego, however, the primary quality of directly surveyed values and to replace them with interpolated ones merely for the sake of cartographic programming convenience seems to us as being opposed to a sound and honest professional surveying and cartographic practice. We will therefore not relate to the problem of the "interpolation" or "computation" of DTM's. Models of this kind must be surveyed, not computed. The earth's surface, being completely irregular, is prima facie, void of any mathematical characteristics. Of course, surveyed values are afflicted with surveying errors which have to be dealt with. But this is unrelated to the use of algorithms of cartographic automation. Surveyed height points are being corrected for a number of detrimental influences such as systematic errors of instruments, influence of the earth's curvature, terrestrial refraction, etc., etc. These corrections are, however, an integral part of the surveying process. The cartographic treatment, both conventional and automated, begins after these corrections have been applied by methods dictated by the surveying profession. If the surveyed DTM is not good enough, the only way to improve it is the way back into the field or to the photogrammetric measuring instrument. Mathematical interpolation cannot be a substitute for professional surveying.

The problem of contour presentation of thematic phenomena is an entirely different matter. The surfaces to be depicted aim to demonstrate the spatial distribution of what is mostly statistical data or, in the case of geophysical or geological maps, subterrainous information. The density of the point values measured is mostly, and especially in comparison with digital terrain models, extremely sparse. A densification is often a very costly matter (more bore holes!) and also the selection of the points is done without previous knowledge of the surface to be mapped as these are statistical and/or invisible surfaces. It can therefore be said that surfaces of this kind are usually very much underdetermined.

In all these cases the "true" surface to be mapped is unknown and the contouring process is mainly an interpretative tool, helping to analyse the phenomenon mapped. As a realistic, mechanistic interpretation alone is obviously impossible, the investigator's judgment, and sometimes bias, is often essential in arriving at a meaningful interpretation and in producing what may be called "reasonable contours." Furthermore, the often expressed wish to achieve "smooth" contours belongs to this category of input. (The word "smooth" should not be confused with "continuous." For topographical contours of the earth's surface, the term "smooth contours" is misleading. These contours are as "smooth" or as "rough" as dictated by the terrain and the map scale. They should however always be continuous except for places of relief discontinuities, such as edges of cliffs, etc.) Most of the DEM's on which such contours are based are of type D. Their character being what it is, there is of course no compelling reason not to transform the geometrical disposition of the random point distribution into a regular array of an interpolated, secondary DEM. As mentioned before, the computer assisted cartographic treatment of a regular array is much more convenient than any other point disposition. As long as the surface depicted by the interpolated grid passes through the usually too few random points measured, any interpolation algorithm will do. There is no unbiased way to judge the degree of approximation of the interpolated surface to the "true", alas unknown, surface. Only later control measurements compared with the values of the interpolated surface can tell how good the computed DEM has proved to be. There are a great number of possible interpolation algorithms and many publications in the field of automated cartography are dedicated to this problem. In our opinion much too much attention has been given to this aspect of computer assisted contouring. As stated before, for surveyed digital models of the earth's surface, the original point distribution should be left unchanged and the contouring software be adapted to the reality of the DTM type at hand. For thematic phenomena, the true surface to be mapped being unknown and/or invisible, there is no way to say which of the purely mechanistic interpolation techniques are to be preferred. The transformation of a DEM of type D to any of the other types is legitimate with any algorithm as long as the interpolated surface passes through the measured points. Computing efficiency may, of course, influence preferences for certain algorithms.

So far we have referred to the contouring of basic maps, i.e., maps for whose production initially surveyed values are used. Contours on secondary maps, which are usually of smaller scale, are, as a rule, the result of the generalisation of the altimetric contents of larger scale maps. In conventional cartography this process of generalisation is applied to the contours of the source maps. If a greater vertical interval is dictated by the reduced scale, a selection from the original contours is made and their flow is adapted to the requirements of the smaller scale. Forms considered to be too minute are eliminated by a "smoothing" of the contours, incorporating them into more prominent relief features. This simplification process of the contours, when done manually, takes into consideration not only the individual contour, but judges its flow in relation to the neighbouring contours and contour configurations. This is basically a geomorphological evaluation process of the relief and its adaptation to the reduced scale to the best of the individual judgement of the cartographer. This is a very subjective activity - its results being, that different generalised maps of the same terrain - even if taken from the same source maps and at the same scale - by different cartographers will never be alike. Applications of computer assistance have tried to imitate the manual method by what has become to be known as "smoothing techniques". These, too, seem to us as being problematic procedures which have to be applied with utmost prudence.

As mentioned before "smooth" should not be confused with "continuous". The linear connection of the points of a contour line threaded into the lattice of a DTM would not result in a continuous curve, the lattice usually not being dense enough. The interpolation of a continuous curve through these points is therefore called for. As long as this curve does not transgress the positional accuracy required from contours of the specific cartographic job at hand, this interpolation is perfectly legitimate. Furthermore small irregularities or bulges caused perhaps by the contour threading algorithm (and not by the geomorphological pecularities of the terrain!) may of course be removed by what may be called a burnishing algorithm. The computer assisted "smoothing", however, of individual contours for the sake of cartographic generalisation is doubtful. An individual contour bears no information about form. It is only in association with neighbouring contours that conclusions about topographical relief can be drawn. This is, in fact, the way that manual cartographers reach their generalisation decisions. We cannot expect that a machine will do a better or even an equivalent job if we supply it with less information.

The best and perhaps the only reasonable way to generalise relief with computer assistance is to apply the generalisation procedure to the initial DTM. This can be done by reducing the number of points contained in the original DTM in such a way that the new surface defines a simplified-generalised relief in accordance with the requirements of the reduced map scale. This process may, of course, change the type of the original DTM (see fig.1) which is one more reason for the need of contouring programs suited to the specific DTM type at hand. It should also be stressed, that the generalisation of the initial DTM should be dictated by geomorphological considerations and not by abstract mathematical techniques. As above-mentioned, topographical relief of the earth surface is void of any mathematical regularity.
The interpolation of contours into such a DTM will automatically result in generalised contours.

The consequence of the above is clearly, that for the computer assisted production of cartographic contours, contouring programs tailored to the specific character of the initial input and the accuracy to be achieved must be employed. The use of so-called "general purpose" contouring programs is therefore, in our opinion, a rather simplistic approach. The problems of cartographic relief depictions of high quality are much too complex to be resolvable by "patent" solutions.

A MATHEMATICAL EVALUATION OF SIMPLIFICATION ALGORITHMS

Robert B. McMaster
Department of Geography
University of Kansas
Lawrence, Kansas 66045

ABSTRACT

A major objective in computer cartography has been the creation of accurate representations of naturally-occurring lines such as rivers, coastlines, and contours. The coordinate data to represent these lines is commonly gathered with digitizers. As a result, the cartographic community has developed an array of line simplification algorithms to eliminate the superfluous data captured when digitizing. Some of the algorithms, such as the nth point and the random point selection routines, have not been developed using sound cartographic principles while others, such as the Douglas-Peucker routine, appear more cartographically sound. By applying two measures of displacement between a digitized line and its simplification, four algorithms have been evaluated. The results indicate that, of the four, the Douglas-Peucker routine produces less displacement between a line and its simplification.

INTRODUCTION

The cartographic community has developed an array of line simplification algorithms over the past decade. Some of these apparently are not based on reasonable cartographic principles, while others appear more sound. Numerous mapping establishments have been working on the problem of linear simplification, but thus far no organized program of research has been published. The needs of the cartographic community would be best served by a comprehensive research design including the following efforts:

(1) The need to measure the various attributes of lines so that comparisons may be made among the different representations of the same line, each produced with differing numbers of coordinates. These attributes include: sinuosity (the curviness or wiggliness) of the line, total length of the simplified line, the linear displacement between a line and its simplification, as well as other geometric measurements.

(2) The need to know which simplification algorithms provide the most accurate representations and, at the same time, eliminate the greatest number of unnecessary coordinates. The analytical process can best be accomplished through the utilization of the measurement techniques developed in (1).

(3) The need to know which algorithms produce the most acceptable simplification from the map user's point of view. Psychological testing can be applied to determine the

threshold limits of user's abilities to detect differences with different degrees of simplification (Jenks, 1978; White, 1983).

(4) The need to set parameters of the algorithms so that a predictable simplification can be generated, one having both the desired perceptual characteristics and the desired storage characteristics. Here, the results obtained in (1) can be associated with the results of (3) to see how the mathematical measurement of lines is related to perceptual observations.

The following paper addresses the first two of these research efforts. First, a method is presented for evaluating line simplification algorithms using two measurements. And, second, four sample algorithms are evaluated.

MEASUREMENTS

The first step in this research was the development of specific measures to evaluate differences between a line and its simplification. Intrinsic to the simplification process is a geometric displacement of the line from the original position. This results in sections of the simplified line no longer lying in the same geographic position as those sections of the original line, or a shift in the true geographic location. But, is this movement consistent along the length of the line or is it highly variable ? The significance of measuring these displacements can not be overemphasized. The most desirable simplification algorithms produce as few of these displacements as possible and, of even greater significance, space these displacement errors evenly along the line. Consequently, the increasing amount of error between a base line and the simplification is most probably the most important indication of the geographic quality of an algorithm. This displacement may be calculated using two different measurements: vector displacement and areal displacement.

Each time part of the simplified line is displaced, as a result of coordinate elimination, a vector difference is produced (Figure 1). This difference may be measured as the perpendicular distance from the eliminated coordinate on the base line to the new vector on the simplified line. For instance, the enlargement of points 1, 2, and 3 illustrate the vector displacement. The distance is measured from point 2, on the base line, to 2B, the perpendicularly projected point on the simplified segment.

A second comparative measure evaluates the areal, or polygonal, difference between a base line and its simplification. This is illustrated on Figure 2. These polygonal differences, as calculated utilizing the formulas developed by Robert Bachi, were summed in several ways (Bachi, 1973). Figure 2 illustrates the positive differences (polygons with a lighter shade on the left hand side of the base line) and negative differences (polygons with a darker shade on the right hand side of the line).

Figure 3 is a graphic comparison of two simplification algorithms highlighting the areal displacement. Using a twen-

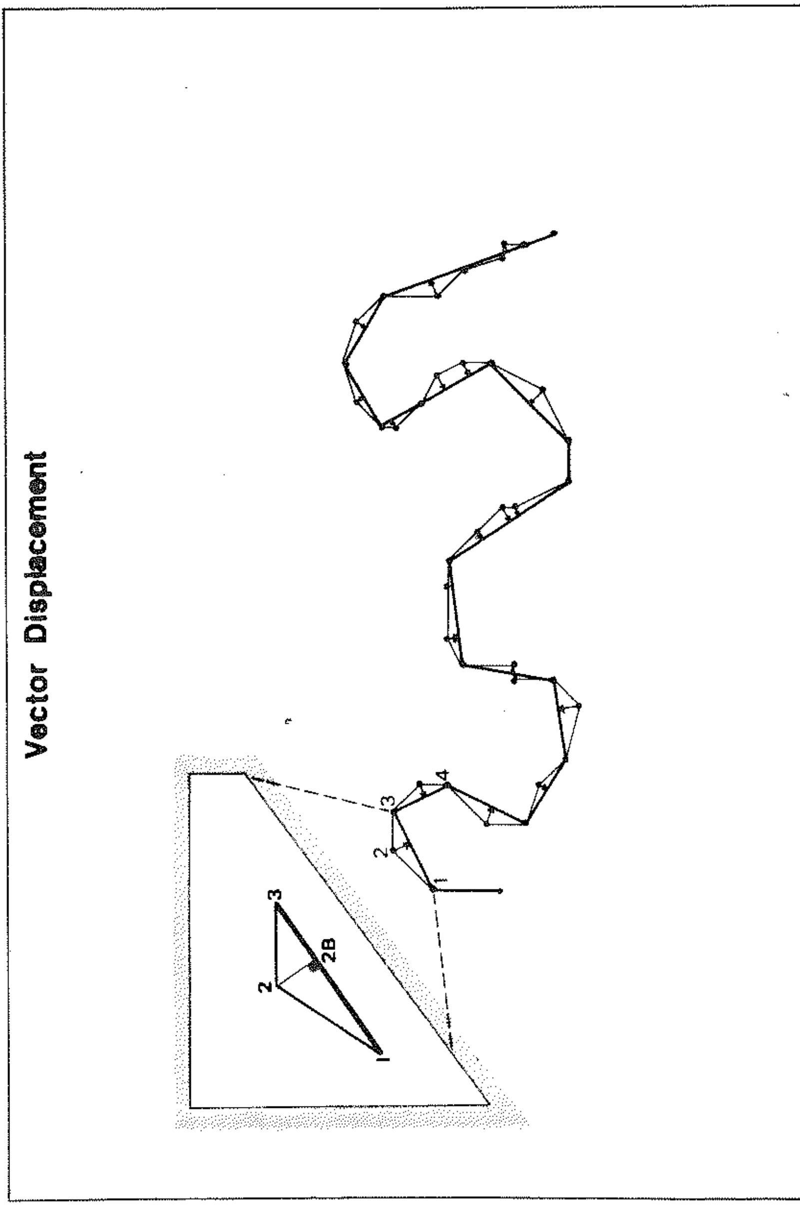


Figure 1. Vector displacement

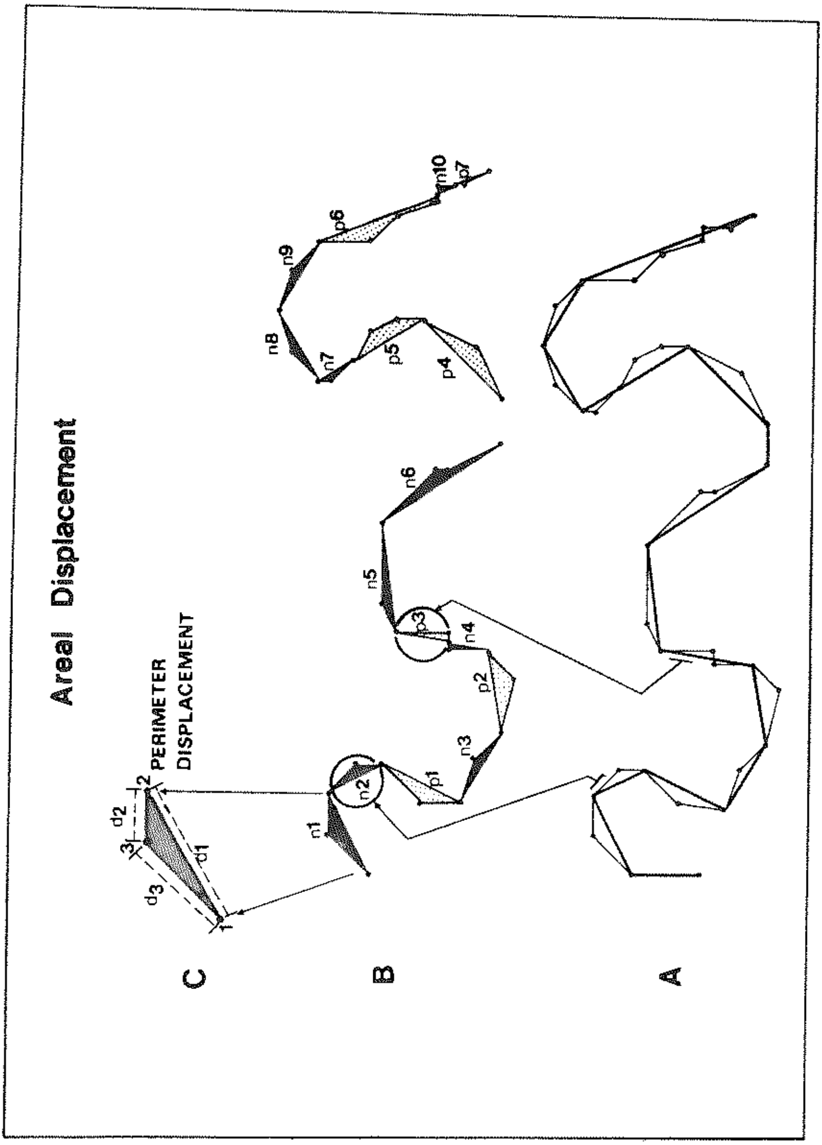


Figure 2. Areal displacement

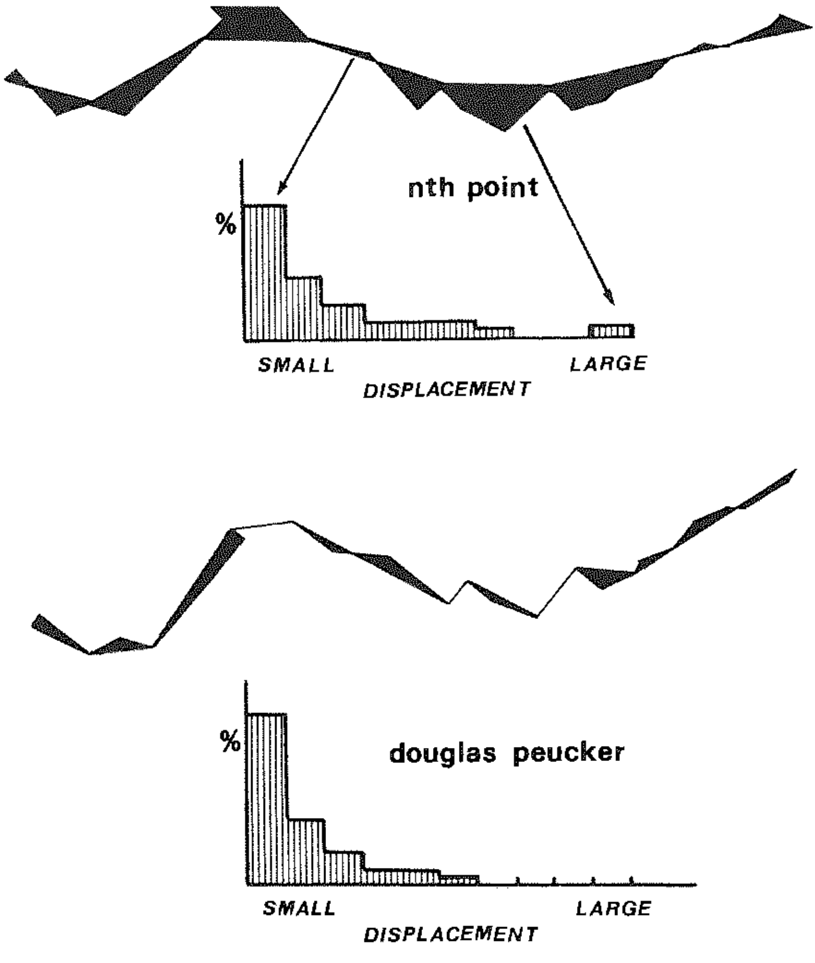


Figure 3. Areal displacement for two algorithms

ty times enlargement of the line, the nth point algorithm is plotted at the top of the page. Note that the displacement, as shown as the black area between the lines, is much less for the Douglas-Peucker algorithm. Simple areal histograms have been plotted for each of the algorithms. The Douglas-Peucker algorithm has more of the smaller and less of the larger, more perceptually significant, displacements. On the other hand, the graph of nth point displacement illustrates a greater number of the larger differences.

VECTOR AND AREAL INDICES

Each of the areal displacements between a line and its simplification was classified into one of ten possible categories based on its size. A very small area would be summed into the smallest category or "bin". A very large displacement, on the other hand, would be summed into category 10. Figure 4 illustrates a series of these histograms. Each histogram has ten size categories on the x-axis, with 1 representing a very small displacement between a line and its simplified version. The y-axis displays the percentage of displacement that fell into each of the size categories. Each of the ten histograms displayed in the figure represents a different degree of simplification. The first histogram (A) is based on a simplification with 656 of the original 889 coordinate pairs. The last histogram (J) is calculated using only 5 coordinate pairs. It is clear that as the tolerance increases, the size of the displacement gets larger. Histogram (A) shows 100 % of the displacement falling into the smallest category (1). Small displacements indicate that the simplified line has not shifted significantly from the original. As exhibited on histogram (E) with 230 coordinate pairs, some of the displacement now appears in the largest category. Most displacement falls into this largest category on histogram (H) with 125 coordinate pairs and, with a severe simplification (J), 100 % of the displacement falls into category 10. Large displacements, of course, indicate a significant shift of the simplified line from the original.

A comparison of these histograms may be quantified by calculating a histogram index. An index is computed in the following manner:

(% displacement in class 1 X 1.0) + (% displacement in class 2 X 2.0) + ...
+ (% displacement in 10 X 10.0)

$$\text{index} = \sum_{i=1}^{10} (\%\ \text{displacement bin } i)\times(i) \qquad (1)$$

Using this method the percentage displacement in each class is weighted by the size of the displacement in each class. The maximum displacement possible is 100 % in category 10, or 1000. By dividing each of the summations by 1000, it is possible to obtain an index value for each of the histograms. An index value of .1 is calculated from histogram (A). As previously mentioned, with increasing simplifica-

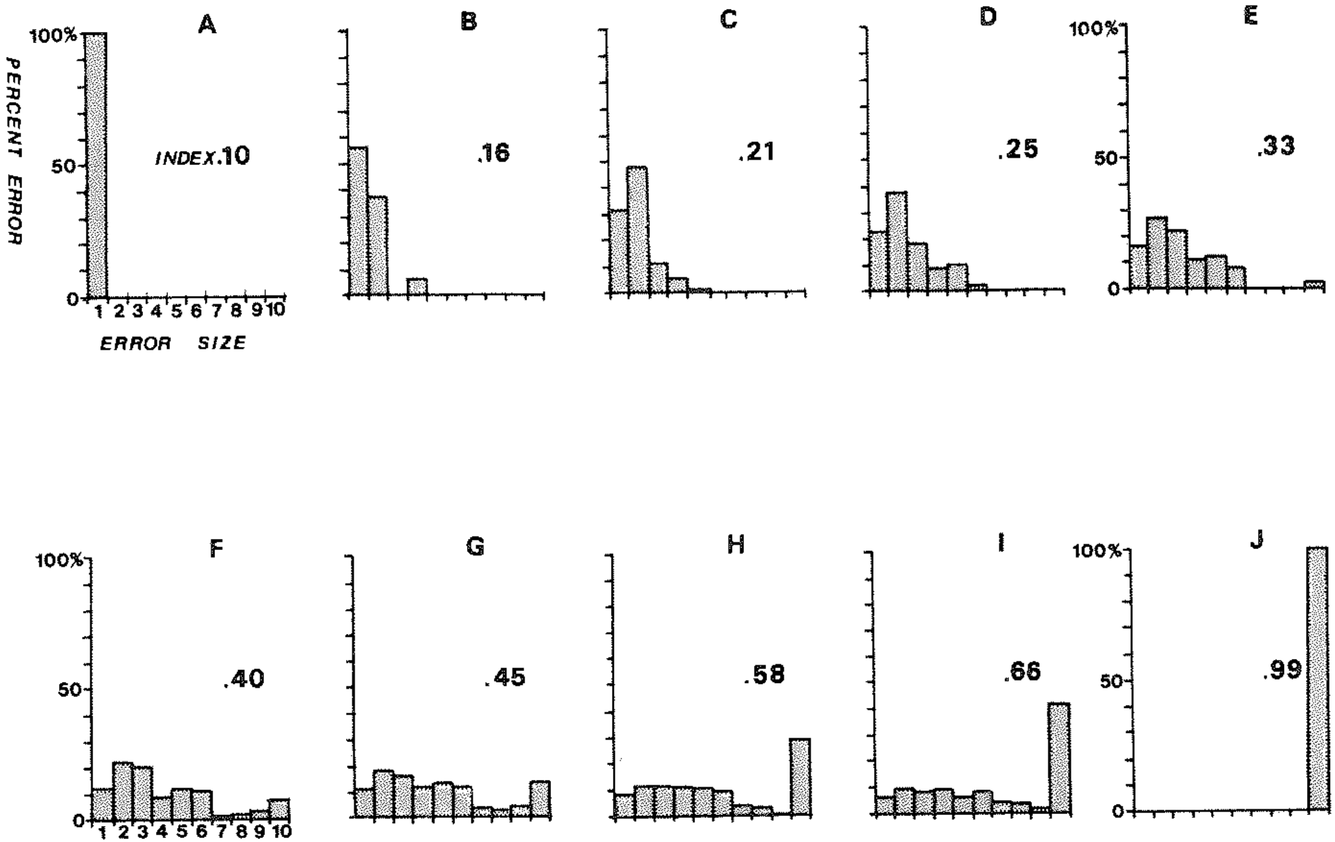


Figure 4. Areal Displacement Histograms

tion the size of the displacements gets larger--as noted by the continual rightward shift of the displacements on the x-axis--producing a final histogram with an index of .99 (J). By using both the areal and vector index it is possible to assign a single number in comparing these histograms.

SIMPLIFICATION ALGORITHMS

Four commonly used algorithms were reviewed utilizing the indices (both areal and vector). One algorithm was the simple nth point routine where every nth coordinate pair is selected as a coordinate pair on the simplified line. A second algorithm used an angular tolerance for eliminating coordinates. The angle between a vector connecting the first and third points and a vector connecting the first and second points is calculated. If this angle is less than a preset tolerance, point 2 is rejected. A third algorithm uses a perpendicular distance calculation in a similar manner. Using a triad of points, a line is calculated between the first and third coordinate pairs. The perpendicular distance is computed from this line to the second point. Once again, if this distance is smaller than the preset tolerance, the second point is rejected. A fourth algorithm was the Douglas-Peucker routine as commonly reported in the literature (Tobler, 1965; Rhind, 1973; Douglas and Peucker, 1973). See Figure 5.

RESULTS

Figure 6 presents a summary of the indices as applied to five digitized lines. The areal index is presented at the top of the page, the vector index at the bottom. Along the y-axes are plotted the index values with the percentage coordinates eliminated plotted on the x-axes. It is clear that, irregardless of the degree of simplification, the Douglas-Peucker algorithm has both less areal and vector displacement. The angular simplification algorithm, on the other hand, has more areal displacement than the other three at all simplification levels and more vector displacement at most levels. Conclusions regarding the vector simplification and nth point simplification are more difficult to formulate. The vector algorithm is clearly superior using either index to a level of approximately 60-70 % of the coordinates eliminated. It is surprizing that, with more rigorous simplification, the graphed lines are coincident indicating an eqivalent amount of displacement. A general ranking of the algorithms' quality would place Douglas-Peucker first, followed by the vector, nth point, and angular algorithms.

CONCLUSIONS

By applying two measures of displacement between a digitized line and its simplification, four algorithms have been evaluated. The two measures included a vector and an areal displacement. Based on these measures two indices were derived from the distribution of displacement sizes. A low index value indicates most of the displacement falls into the smallest category while a higher index value will

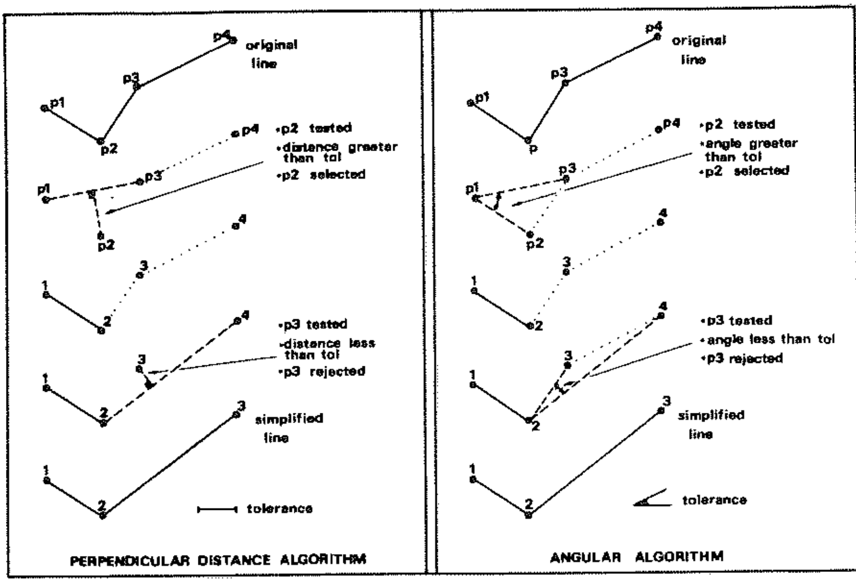


Figure 5. Vector and angular simplification algorithms

result from large displacement. Using these indices on the results of four simplifications, it appears that the Douglas-Peucker routine produces less displacement between a line and its simplification.

ACKNOWLEDGMENTS

The author wishes to thank Mark Lindberg of the Department of Geography, University of Kansas for his assistance.

REFERENCES

Douglas, D.H. and Peucker, T.K. 1973, Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature: The Canadian Cartographer, Vol. 8, No. 2, pp. 112-122

Jenks, G.F. 1978, Perceptual Thresholds in Linear Generalization by Computer: Association of American Geographers Program Abstracts, pp. 229-230, Washington D.C., Association of American Geographers

Rhind, D.W. 1973, Generalization and Realism Within Automated Cartographic Systems: The Canadian Cartographer, Vol. 10, No. 1, pp. 51-62.

Tobler, W. 1965, Automation in the Preparation of Thematic Maps: The Cartographic Journal, Vol. 2, No. 1, pp. 32-38.

White, E.R. 1983, Perceptual Evaluation of Line Generalization Algorithms, M.A. thesis, Virginia Polytechnic Inst.

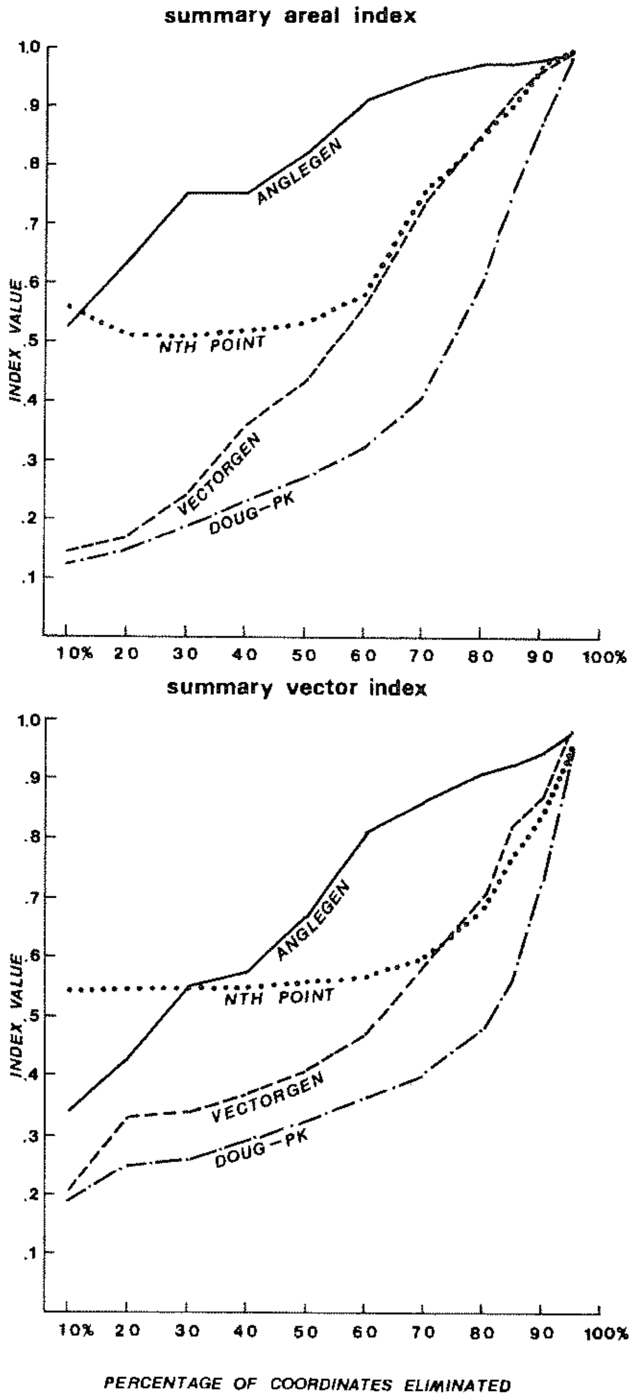


Figure 6. Summary areal and vector indices

VISUAL VERSUS COMPUTERIZED SERIATION: THE IMPLICATIONS FOR AUTOMATED MAP GENERALIZATION

J.-C. Muller
John L. Honsaker
Department of Geography
The University of Alberta
Edmonton, Alberta T6G 2H4

ABSTRACT

Seriation as a classification procedure has been used by various disciplines, including archeology, linguistics and geography. It can be used also as a generalization procedure for categorical and non-categorical map data. Two major problems have limited the wide acceptance of geographic seriation for mapping purposes: 1) the absence of objective criteria to recognize successful seriations and 2) the enormous amount of time required to rearrange geographical administrative units by trial and error to find the geographical sets characterized by similar attributes. Objective criteria for seriation are proposed and a heuristic automated seriation procedure is presented. It is shown that visual seriation is unreliable, unpredictable and requires much time and professional experience. It is suggested that geographic seriation is a non deterministic process for which there is no perfect mathematical solution. Instead, automated seriation must be considered as a working solution leading to multiple alternatives to the problem of map generalization. It provides the necessary background for further analysis based on visual methods and geographical thinking.

Consider the following problem: one wishes to regionalize Canada into groups of provinces which show similar economical structures. Assume that we are given the percentages of the provincial total labor forces involved in nine distinct types of occupations for the eleven provinces. One could graphically transcribe the various statistics in a matrix form according to the following binary code : a matrix entry is painted black if, for a particular province (column), the corresponding occupation (row) mobilizes a percentage of the provincial labor force which is above the national Canadian average; otherwise it remains blank. One could further rearrange the columns so that the position of each province best reflects the degree of similarity between that province and all other provinces in Canada. Likewise, the rows could be rearranged in such a way that labor force types which are commonly shared by the same provinces are regrouped together. A partitioning of the matrix along the columns into subsets of provinces characterized by similar dominant labor forces could provide the basis for a spatial regionalization of Canada and a map symbolizing the regional structure

(Figure 1).* Whether this type of regionalization is geographically or economically the most relevant is probably questionable. It is apparent, however, that the type of classification described above is informative and could help geographical thinking.

Seriation as a classification procedure has been used for quite some time by a variety of disciplines, including ethnography, linguistics, psychology, physical anthropology, archeology and, later, biology and geography.** Archaeologists emphasize more the sequencing aspect of the technique to solve problems of chronology between various artifacts, whereas biologists have used seriation in connection with numerical taxonomy. One of the most notorious applications has been performed in relation to spatial organization and regionalization. Bertin, a leading French cartographer, has developed a school of thought in graphic information processing which is strongly based on seriation (Jacques Bertin, 1977). He has added a spatial dimension to the seriation methods by expressing the results in a map form, as shown in Figure 1.

There has been two distinct problems related to seriation studies: 1) one must define suitable criteria to recognize a successful seriation among several trial orderings of an item set and 2) there is the enormous amount of time necessary to rearrange the item positions by trial and error to find the best seriation of items. The two problems are obviously interconnected. Particularly, one needs an objective definition in order to implement an algorithm which could speed up the process of seriation through automation. One purpose of this paper is to show that visual seriation alone, unsupported by automated procedures, leads to widely variable results. Some automated alternative solutions will be considered confronting more particularly exact versus heuristic methods. As a matter of introduction, only binary relations will be considered in this study, although the methods presented are clearly extensable to more sophisticated metrics.

VISUAL SERIATION

The Canada labor force unseriated matrix was presented to twenty four University students who had had previous training on visual seriation. They were explained that the objective of seriation is to bring together rows and columns which are most alike. They were to follow the following manual procedure: 1) take one row at random, and move it to the top; 2) permute the columns so that

---

* The data are found in Statistics Canada, Employment, Volume reference: 71-001, Ottawa, 1971. Each employment value was expressed as a percentage of total employment within the specific province to suppress the effects of the vastly different employment totals.

** See an extensive bibliography of applications in Leroy Johnson, Jr. Item Seriation as an Aid for Elementary Scale and Cluster Analysis (Eugene: University of Oregon, Bulletin No. 15, 1968), p. 2.

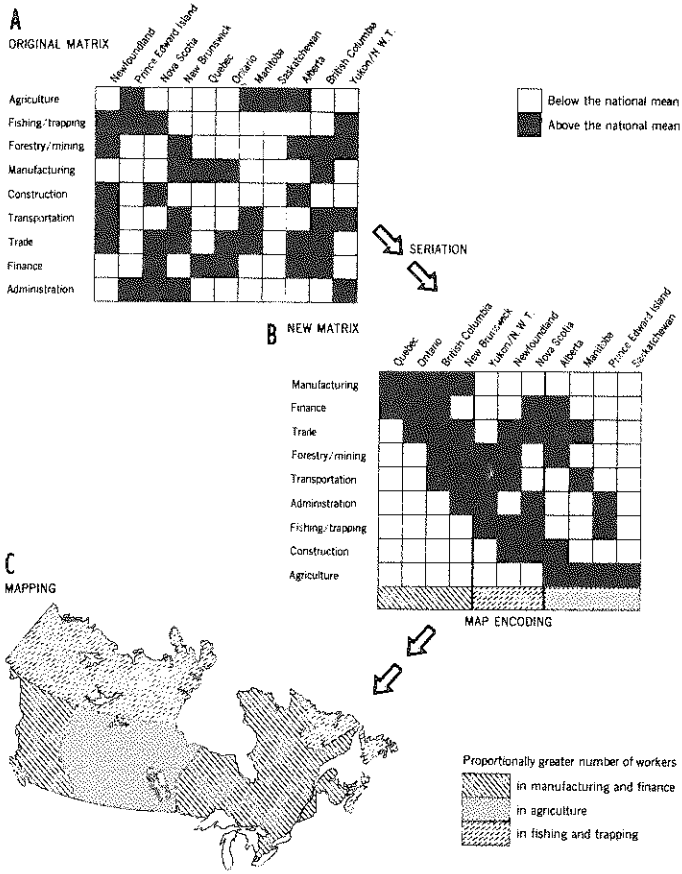


Figure 1. A Regionalization of Canada Based on the Seriation of the Relationships Between Provinces and Labor Force Percentages.

the top row has a "regular profile" (in this case all black cells occur consecutively); 3) permute the rows so that those most "similar in profile" to the top row are placed next to the top, the most dissimilar or "opposite in profile" on the bottom of the matrix; 4) leave in the center of the matrix the rows which bear no relationship with the top row and reorder the center section according to row similarities. The subjects were able to permute rows and columns using a device developed by Bertin, whereby the matrix is made up of individual dominoes with holes on four sides so that thin steel rods may be passed through an entire column or an entire row of the matrix. The test was conducted on "blank" matrices. Rows and columns were only identified by numbers in a random sequence, so that no bias would be introduced by the subjects' preconceived ideas about spatial economy in Canada. The results raise two questions: 1) How much agreement there is between the responses, and 2) what is the quality of the seriation performed by the subjects.

One obvious mean to appreciate the degree of congruence between the responses is to stack up the twenty four matrices on top of each other and measure the cumulated sum $C_{ij} = \Sigma\, a_{ijk} = \{a_{ijk} | a_{ij} = 1\}$, k = 1, 24 for every i and j.* Total agreement would be indicated by a matrix C displaying only two possible values for the matrix entries: either $C_{ij} = 24$ or $C_{ij} - 0$. A departure from those extremes indicate some disagreement. Maximum disagreement occurs when the values C_{ij} lay somewhere around the midpoint between the extremes. Perusal of the cumulation matrix C shows overall little agreement between the responses.

A more refined approach to measure the degree of concordance between the responses is to calculate the mode response matrix. The dominant location for each province and each labor force is calculated. A percentage value is associated to each location describing the number of supporting cases among the twenty four respondents for that particular location (Figure 2). Again, only few rows (Trade, Forestry/Mining, Transportation, Construction) and few columns (Saskatchewan, Quebec) are supported by a majority in their mode location. All other mode locations represent the views of a minority, indicating little agreement between the subjects tested. A third measure of disagreement can be introduced by measuring the structural distance between the response matrices. Structural distance is defined by the number of swaps necessary to introduce in the rows and columns of matrix A in order to get matrix B. Thus, TA = Ø iff A ≡ B. Consider the two following sequences describing the row orders of matrix A and B. Four swaps are necessary in order to go from A to B:

B	1	2	3	4	5	6
A	2	6	3	1	4	5
Swap 1	1	6	3	2	4	5
Swap 2	1	2	3	6	4	5
Swap 3	1	2	3	4	6	5
Swap 4	1	2	3	4	5	6

We found empirically that the swap counting is independent from the choice of location in reordering the elements of A according to the order of B. Comparison of the response matrices show that the average number of swaps to go from one matrix to any other is 5.7 on the rows and 7.4 on the columns. These are relatively large numbers considering the small size of the matrices, pointing out a wide variety of responses.

* Matrices were all similarly "oriented" so that no mirror effect would interfere with the amount of disagreement between the subjects in sequencing the rows or the columns.

The outcome of visual seriation appears to depend strongly on the subject being tested. The variance of responses is probably related to differing opinions and aptitudes in judging row and colu similarities. One would expect also the variance to increase as the complexity and the size of the unseriated matrix increase.

AUTOMATED SERIATION

Any attempt at automation of the seriation process requires a mathematical formulation of the objectives. Let $A = (a_{ij})$ be an m x n matrix whose entries a_{ij} may assume one of two values. The matrix A may be regarded as the incidence matrix of elements $e_1, e_2, \ldots, e_m$ versus sets $S_1, S_2, \ldots, S_n$; that is, a_{ij} is equal to 1 or 0, black or white according to whether e_i is or is not a member of S_j.

Consider two rows e_i = 10100 and e_j = 01100. e_i and e_j are distinct points in a fifth-dimensional space. They are vertices on a hypercube. Impose the "city block" metric on this space, so that the distance between rows e_i and e_j is equal to the number of 1s in e_i, not in e_j, plus the number of 1s in e_j not in e_i. This distance is called the Hamming distance in the graph theory literature (Kaufmann, 1970). An optimum sequencing of n rows e_i, $i = 1, \ldots n$ will be obtained when the sum of the Hamming distances between consecutive rows will be minimized. Thus, the problem is identical to finding a Hamiltonian walk in a finite graph, that is the shortest path which connect all vertices once and only once. Each graph is scanned separately, one for the rows and one for the columns. The procedure was implemented on the University of Alberta's Amdahl computer in the form of a Pascal program.* As it turns out, there are many equally optimum solutions to the seriation of the Canada matrix. This implies that there are no deterministic answer to the problem of seriation. The user is given the opportunity to select among the various alternatives, the solution which fits best with his needs.

Unless the matrix is relatively small, the minimum path solution is not practically feasible, however. The computation time is of order $(m-1)! + (n-1)!$, where m and n are the numbers of rows and columns. The seriation of the Canada matrix required 250 seconds CPU time. A n! search of 70 rows would require approximately 10^{68} years with an IBM system/370! One needs to use an heuristic method which would handle bigger size matrices in much shorter computer times.

The major difficulty in using the minimum path method is the necessity to look at every possible sequencing combination of n objects of n attributes, computing successively the length of all the possible paths described by a minimum Spanning Tree. There are sequences which are highly repetitive, however, and which from a computational

* The program is available on request from the Department of Geography, The University of Alberta.

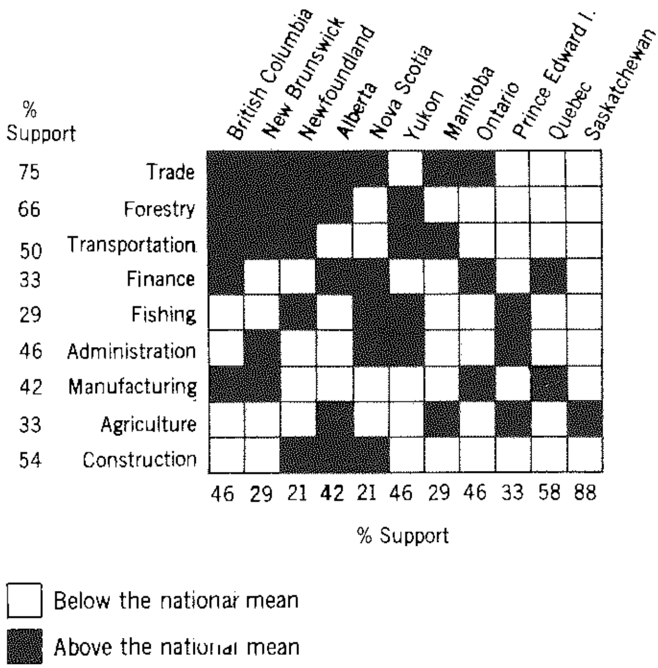


Figure 2. The Mode Response Matrix. Percentage Values Indicate the Proportion of Respondents who Supported the Particular Positions of Provinces (Columns) and Employment Types (Rows) Shown in the Matrix.

as well as a substantive view point lead to identical results. The following sequences of objects {A, B, C, D, E, F}:

A	B	C	D	E	F	F	E	D	C	B	A
B	C	D	E	F	A	E	D	C	B	A	F
C	D	E	F	A	B	D	C	B	A	F	E
D	E	F	A	B	C	C	D	A	F	E	D
E	F	A	B	C	D	B	A	F	E	D	C
F	A	B	C	D	E	A	F	E	D	C	B

give identical path lengths if one goes back to the origin. The path becomes a cycle of objects arranged along a closed belt. One would like to take advantage of this cyclic redundancy. Only those cycles which show different neighbourliness between objects, such as A B C D E F and B A C D E F, would be compared. This idea was exploited by De La Vega in 1972 who developed a seriation algorithm based on successive triad arrangements of clusters (W.F. De La Vega, 1972).

In order to understand the procedure, one has to recall the theorem of Kendall which states that given the row-permutation of a seriated matrix, then the row-permutations which give to A the seriated form are exactly those, which, when applied simultaneously to the rows and columns of $S = AA^T$, give to that the Robinson form (David G. Kendall, 1969). Furthermore, a matrix S has a Robinson form if and only if the columns and the rows are unimodal and attain their maximal values on the principal diagonal. Thus, the elements of a Robinson matrix never increase when going to the left or down from any position on the main diagonal. A Robinson form is only an ideal, however, and most similarity matrices $S = AA^T$ will not attain this perfect shape because the parent matrix A is not perfectly "seriatable". But optimizing the Robinson form on S will optimize the seriation on A.

The first step in applying De La Vega's algorithm is to calculate the clusters on the rows and the columns of A. The similarity coefficients between rows and between columns are derived from S (rows) = AA^T and S (columns) = A^TA. The weighted pair-group average linkage was used to determine the clusters (Peter Sneath and Robert Sokal, 1973). One obtains two dendograms showing a new sequencing between the rows and between the columns after the clustering on S (rows) and S (columns).

The second step is to reorder this sequencing by successively analyzing the similarity between triads of elements belonging to different clusters. This analysis leads to a reordering of S so that it approximates a Robinson form, a condition for the seriation of the parent matrix A.

This cluster/triad combinatorial solution was implemented at the University of Alberta in the form of an APL Program.* The computing time is of order $m^3 + n^3$, where m and n are the numbers of rows and columns. N^3 is much smaller than N! and represents a considerable saving of computation time. An originally N P complete problem has now been converted into a tractable exercise. The question, of course, is whether the results of the heuristic approach are nearly comparable to those of a brute force method such as the shortest path optimal solution.

COMPARISON BETWEEN HEURISTIC AND OPTIMAL SOLUTIONS

In order to compare objectively the performance of the heuristic versus the optimal methods, it is necessary to measure the degree of seriation displayed in the solutions. The Tonnelier measure was already introduced as one possible criterium of success of a seriation solution. Another criterion could be the cumulated distance between rows and between columns, $\Sigma\, d\,(r_i, r_{i+1})$ and $\Sigma\, d\,(c_j, c_{j+1})$. One could also evaluate the quality of the seriation solution through examination of the corresponding similarity matrices on the rows and on the columns. It can be shown

* The program is available on request from the Department of Geography, The University of Alberta.

that if the sum of the terms $\Sigma\, S_{i,\, i-1}$ immediately beneath the principal diagonal of S (rows) is trace (S)-m, then all the 1s on the columns of the parent matrix A occur consecutively (Martin E. Wilkinson, 1970). Similarly, if the sum of the terms $\Sigma\, S_{j,\, j-1}$ of S (columns) is trace (S)-n, then all the 1s on the rows of A occur consecutively. When both conditions are fulfilled, matrix A is said to be Petrie on the rows and on the columns. Seriation on real world data rarely leads to a Petrie form, however. The Petrie ratios $\Sigma\, S_{i,\, i-1}$/trace (S)-m and $\Sigma\, S_{j,\, j-1}$/trace (S)-n provide one mean to measure the degree of success of a seriation solution. Another measure, conceived by Tonnelier, was used to evaluate the goodness of the seriated response matrices (W.F. De la Vega, 1972).

$$D\ (M) = \left(\frac{1}{\Sigma a_{ij}}\ \Sigma\ (j-i)^2\ a_{i,j}\right)^{\frac{1}{2}}$$

D (M) increases as more non-zero terms are located away from the diagonal.

A calculation of the Tonnelier and Petrie ratios on the non-seriated, heuristically and optimally seriated versions of the Canada Labor Force matrix shows that the heuristic algorithm provides a nearly optimal solution (Figure 3):

Canada matrices	Tonnelier index	Petrie ratios	
		Rows	Columns
Unseriated	4.32	0.47	0.53
Heuristic solution	2.50	0.69	0.80
Optimal solution	2.50	0.72	0.83

Note that the Tonnelier measure is not as sensitive as the Petrie ratios. Tonnelier is measuring the degree of diagonal patterning whereas the Petrie ratios characterize the degree of contiguity of the 1s elements in the binary matrix. In short, Tonnelier responds more to shape and Petrie more to structure. The results provided by the heuristic method are very satisfactory, specially if one takes in consideration the computing resources utilized : the optimal solution required 250 seconds CPU time; the heuristic solution took only one second, clustering analysis included!

PERSPECTIVES

A recent debate between proponents of visual versus automated seriation shows that automated solutions are a point of departure in the process of seriation, not a point of arrival (Jacques Bertin, 1981). Visual seriation comes after automated seriation, once all possible équally optimum solutions have been provided. Ultimately, it is left to the user to decide which subject belongs to which group. The researcher controls and orients the classification suggested by automated seriation according to his own intuition, imagination and scientific knowledge of the problem being treated. This philosophy has one important implication for cartography. A map is a generalized picture of reality and there are many generalized

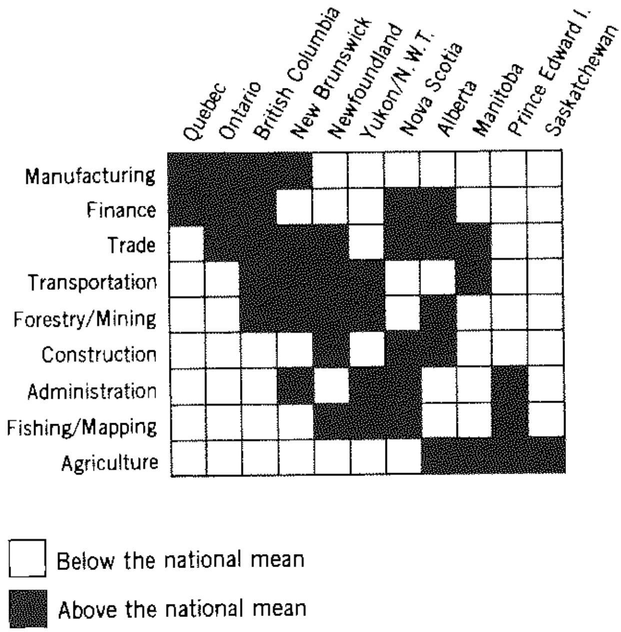


Figure 3. The Seriated Matrix According to the Heuristic Solution. Compare with the Optimum Solution of Figure 1.

options available. Generalizations based upon geographic seriation are not unique and it is left to reader to choose among all possible maps the one which suits best his needs. Thus, geographic seriation emphasizes the view that cartographic generalization is not a deterministic process.

Since automated seriation appears to be a first step in the development of a visual solution, one would like to use an algorithm which provides a seriated pattern closely related to what is thought to be a visually acceptable answer. One also wonders if the visual solutions generally conform to the criteria developed earlier for automated seriation. An experiment was conducted with seriated data already available (Henri Leredde and Patrick Perin, 1980). 59 archeological artifacts found in Northeastern France and dated from the sixth century are characterized by 26 attributes ranging from techniques of fabrication to shape and decoration. The relationship between objects and attributes is encoded into a 59 x 26 binary matrix. Among the solutions provided are two automated seriations based on Factor analysis and Cluster analysis and a visual seriation. We have added to those one automated solution derived from the De La Vega heuristic algorithm presented in this paper (Figure 4). A calculation of the Robinson index and Petrie ratios on the various matrices show that the De La Vega method provides a solution at least as

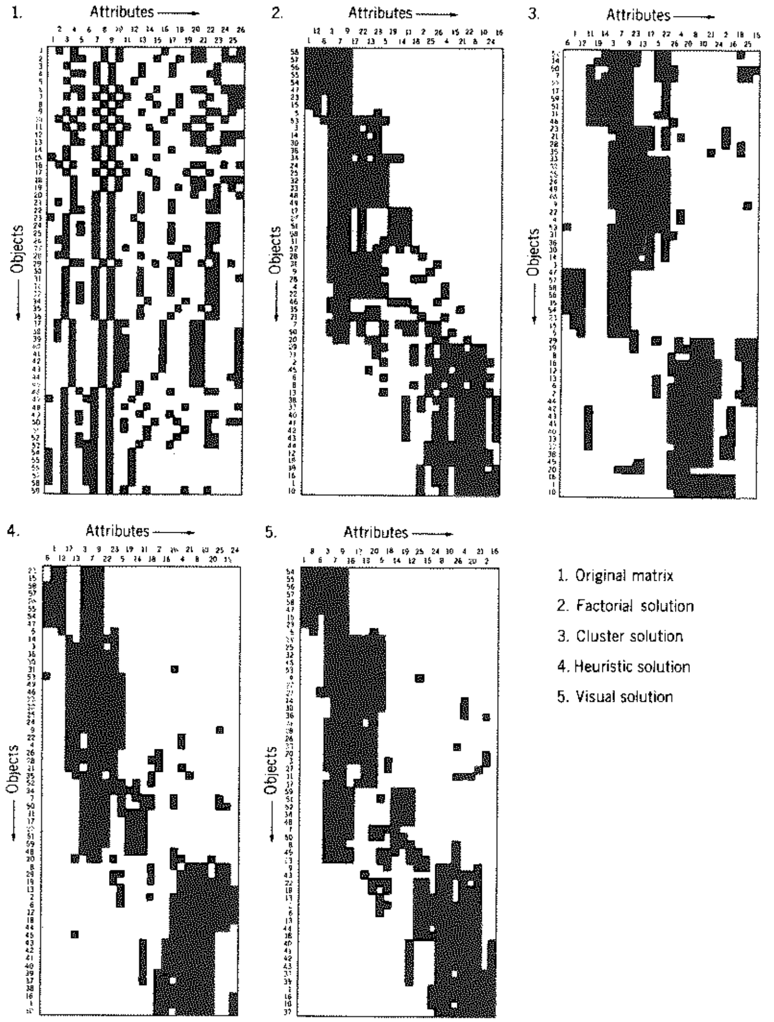


Figure 4. The Heuristic (De La Vega) Solution is as Good as the Visual Solution, and Better than the Cluster and Factorial Solutions.

good as the visual seriation:

Matrices	Tonnelier index	Petrie Ratios	
		Rows	Columns
Unseriated	25.45	0.61	0.17
Factor Analysis	22.33	0.92	0.71
Cluster Analysis	23.08	0.97	0.74
De La Vega's method	18.30	0.99	0.75
Visual method	22.37	0.98	0.77

Factor Analysis gives the worst results. Although it is difficult to predict the outcomes of visual seriation, this small experiment shows that there is a clear relationship between the visual method and the objective criteria developed for automated seriation.

Seriation on binary or non-binary data is a form of analysis which has been somewhat neglected, particularly in geography. Geographers were somewhat reluctant in using a concept whose theoretical foundations are not as firm as those from which other multivariate statistical methods were developed, such as Factor Analysis or Muldimensional Scaling. Yet seriation is not just another method of classification and could be used advantageously in the automated generalization of categorical or non-categorical map data. We have shown one heuristic automated solution which lead to results highly comparable to those of visual seriation. Other more traditional tools based on multivariate statistics do not appear to perform as well. Additionally, seriation is a more descriptive approach which does not require extensive technical knowledge in order to understand and critisize the results of the analysis. More fundamentally, the numerical treatment is only a limited part of the classification process, and the reader is free to choose among all possible alternative seriation solutions those which deserve additional treatment. Extensive opportunity for interaction between mathematical and subjective classifications is perhaps the most original feature offered by seriation.

REFERENCES

Bertin, J. 1977, La Graphique et le Traitement Graphique de l'Information, Paris: Flamararion.

De La Vega, W.F. 1972, "Deux Algorithmes de Seriation" in Les Methodes Mathematiques de l'Archeologie, C.N.R.S., pp. 167-184.

Kaufmann. 1970, Des Points et Des Fleches, Paris: Dunod, p. 131.

Kendall, D.G. 1969, Incidence Matrices, Interval Graphs and Seriation in Archeology," Pacific Journal of Mathematics, 28, pp. 564-570.

Sneath, P. and Robert Sokal. 1973, Numerical Taxonomy, San Francisco: W.H. Freeman and Company.

Wilkinson, M.E. 1970, "Archeological Seriation and the Travelling Salesman Problem." In Mathematics in the Archeological and Historical Sciences, Edinburgh: The University Press, pp. 276-283.

Leredde, H. and Patrick Perin. 1980, "Les Plaques - Boucles Merovingiennes," Dossiers de d'Archeologie, Vol. 42, pp. 83-87.

AUTOMATED DETECTION OF DRAINAGE NETWORKS FROM DIGITAL ELEVATION MODELS

David M. Mark
Department of Geography
State University of New York at Buffalo
Amherst, New York 14260, U.S.A.

ABSTRACT

Two algorithm for the automated detection of drainage networks from digital elevation models (DEMs) are described and evalusted. One of these detects points of local upward concavity, while the other simulates runoff processes to predict channel locations. The former algorithm is much faster, but is abritrary, and is only weakly related to hydrologic processes. The latter algorithm is physically sound, but very costly. Both algorithms have difficulty with DEMs containing pits due to data errors; these problems are alleviated by filtering or smoothing the data before channels are detected. Results and evaluations are presented.

The topology and geometry of drainage networks has been an important area of study within theoretical geomorphology for almost four decades. Furthermore, the drainage network is very important in models of landscape development, as well as in drainage basin hydrology. Drainage networks can also form a vital element in geographic information systems for land systems analysis and resource management.

One problem in drainage networks studies is the amount of effort which is needed to identify the drainage net and to measure basic properties. First, the drainage network must be delimited on maps or aerial photographs, or surveyed in the field; then, link lengths, junction angles, and other properties of interest must be measured. During the last two decades, the measurement phase has been simplified and its time demands reduced through the use of coordinate digitizers (see Jarvis, 1977). Once the drainage net has been identified on the source documents, the drainage network can be traced and the coordinates encoded. From these coordinate files, link properties can readily be computed. While this represents an improvement over manual measurements, the digitization phase can still be very tedious, and furthermore, the channel network must still be identified.

Recently, there has been a great increase in the quantity of topographic and hydrographic data available in machine-readable form. Notable in North America are the Digital Elevation Models (DEMs) and Digital Line Graphs (DLGs) currently being produced and distributed by the United States Geological Survey (Elassel, 1978; Allder and others, 1982). If these data resources can be used to quantify drainage network measures automatically, research on drainage network topology and geometry would be aided considerably. This paper addresses the problem of automated detection of drainage networks from digital elevation models. Two algorithms for drainage network identification are discussed and illustrated. One is based on the detection of local surface concavity (after Peucker and Douglas, 1975), while the other simulates runoff concentration, with those cell receiving simulated surface runoff above some pre-determined threshold being declared to be the drainage network (after Speight, 1968). The important roles of data acquisition methods and of data filtering are discussed, and the results of the two algoritms are compared with "traditional" manual methods of characterizing the drainage network from maps.

DEFINITION OF THE DRAINAGE NETWORK

The drainage network which is of interest to most geomorphologists is simply a map of all those points at which fluvial processes are sufficiently dominant over slope processes that a channel is maintained. It is generally recognized among geomorphologists that the "blue line" stream networks printed on most North American topographic maps at scales of around 1:24,000 and smaller are very conservative representations of the actual drainage network as defined above. Blue lines are draughted on the maps at publication scale only where permanently flowing streams, or major intermittent or ephemeral streams, are present. Many minor channels which even in a humid area carry water only during precipitation or snowmelt events are omitted from the maps, even though they usually comprise the majority of the total length of channels in a region. Thus it has become a standard procedure for workers to extend the drainage network up any small valleys whose presence is indicated by alligned, concave-downhill cusps or crenulations in the contours.

When compiling a contour map, USGS topographers are instructed to first sketch in all channels, and then to draw the contours so as to show these channels (Mark, 1983). The channels themselves, however, are generally omitted from the blue lines printed on the map unless they are permanent or major features. Thus, when the geomorphologist extends the drainage network to include small crenulation valleys, he or she is in effect restoring valleys detected by the topograpgher and which, in most cases, probably appeared on the compilation sheet.

There have been a few attempts to employ quantitative rules governing the inclusion or exclusion of concavities indicated by contour crenulations (for example, see Lubowe, 1964). By and large, however, most workers have relied upon subjective judgement to identify "definite" valleys. Krumbein and Shreve (1970) found that there was a very high degree of correspondence between the networks identified independently by experienced researchers. While there was much more variance among inexperienced operators, such individuals could be trained relatively quickly. Nevertheless, this subjective element in drainage network research seems undesireable. The acknowledged inadequacy of the printed blue line networks on topographic maps clearly implies that the USGS Digital Line Graphs do not represent a viable data resource for drainage network studies, since they are obtained by digitizing the blue line networks on maps. Thus attempts to automate the data collection phase of drainage network studies must turn to Digital Elevation Models (DEMs) as a data source.

USGS DIGITAL ELEVATION MODELS

Several years ago, the U.S. Geological Survey began to distribute gridded elevation data sets which had been collected from 1:250,000 scale topographic maps by the Defence Mapping Agency. These sets, termed Digital Terrain Models by the survey, cover 1 by 1 degree areas, and have a grid cell size or spacing of 63.5 metres (208 feet) on the ground; however, since the elevations were obtained by interpolation from digitized contours, the effective resolution is not this good. More recently, the USGS has begun to distribute a new series of elevation grids which they term Digital Elevation Models (DEMs). These cover quadrangle areas 7 1/2 by 7 1/2 minutes (the same size as the USGS 1:24,000 scale topographic map series) and have a 30 metre grid size. Unlike the earlier series, these models have not been obtained from maps, but are a by-product of an orthophotomapping program. In ortho-photomapping, elevations must be generated in order to compute and remove parallax effects; the elevation data can, of course, be stored and used for further processing.

Some of the USGS DEMs have been obtained using the Gestalt Photomapper II, or GPM II (Swan and others, 1978; Allam, 1978; Allder and others, 1982). This image correlation device produces grids of elevations at a spacing of 0.182 mm at the scale of the aerial photographs; these grids are then re-sampled to produce a 30 metre (ground scale) grid alligned with the axes of the Universal Transverse Mercator (UTM) coordinate system. The USGS also produces orthophotos using semi-automatic profiling devices; these yield elevations along profiles. Generally, the spacing of points along the profiles is considerably less than 30 m, while the distance between profiles is considerably greater than this value. When these heights are re-sampled to a 30 metre square grid, the interpolation technique produces estimated elevations which have a much stronger spatial autocorrelation along the direction parallel to the profiles than in the orthogonal direction. Many images produced from these models have clearly visible "stripes" parallel with the scanning direction of the profiler. Nevertheless, the over-all root mean square error in these models is considered to be less than the error in models produced using the GPM II. The re-gridded data are reported as integers in metres above the sea level datum, regardless of the data collection method.

The data tapes distributed by the USGS include file headers containing information on the geographical location of the area covered by the DEM, the method used to collect the data, the date and place that this was done, the minimum and maximum elevations, and other characteristic of the models; these data are followed by the elevations, reported as north-south profiles. Kikuchi and others (1982) discussed the analysis and display of these models in a mini-computer environment They found that pre-processing the data and saving the elevations using system-specific block read and write commands produces a considerable imrovement in the time required to analyze and display these data sets.

These USGS DEM data sets have been used for all testing of algoritbms reported in this paper; however, similar results should be obtained for any DEMs collected and re-sampled in the same manner as these.

AUTOMATED DETECTION OF DRAINAGE BASINS: LOCAL SURFACE COMPLEXITY

Any portion of a topographic surface which is locally concave-upward will be a place where surface runoff will tend to be concentrated. Thus a map of the concave-upward portions of a digital elevation model could be considered to be an approximation of the drainage network. David H. Douglas developed an algorithm for identifying such concave-upward portions of a DEM (Peucker and Douglas, 1975). This algorithm simply flags the highest point among each square of four mutually-adjacent points in the grid. Once this flagging phase has been completed, all those points **which have not been flagged** are the estimated channel network. Note that, except for points along the boundary of the DEM, each point is involved in exactly four comparisons.

One of the programs in the Digital Elevation Models Graphic System (Kikuchi and others, 1982) was modified to produce program HILO, so named because it detects local high and low points. This FORTRAN program reads two north-south profiles from the DEM into core, and then applies the algorithm to all sets of four mutually-adjacent points. Profiles are processed from west (left) to east (right), and so at each stage, the unflagged points in the left (westernmost) profile can be plotted, before a new profile is read. The program can also detect ridges simultaneously; in this case, the **lowest** point in each set of four is flagged, and the unflagged points are the ridges. The program can display the ridges only, the channels only, or both.

Program HILO was used to process the DEM representing the topograpby of the Keating Summit, Pennsylvania, quadrangle. The study area is located in the well-dissected Appalachian Plateau region in northwestern Pennsylvania, and has relatively high local relief and coarse topographic texture. The alogorithms should perform well for such an area. Figure 1 shows the results, which are far from ideal. The drainage network has clearly been identified, but there are also numerous isolated pixels, moat of which repreaent pits or closed depressions. These pits are concentrated in the relatively flat areas, the valley floors and the undissected plateau remnants. Since pits detectable at a resolution of 30 metres are essentially absent in fluvially eroded topography (except for limestone areas), these pits presumably represent "errors" in the models. The pits appear only in the flatter areas because of what is essentially a signal-to-noise ratio effect; while the pits are numerous, most are only one or two metres deep (see Table 1). Where slopes are gentle, a 1-metre negative error may be suffiecient to produce a closed depression, while on steeper slopes, a pit does not result.

The Keating Summit DEM was produced using a GPM II. It is clear from a visual inspection of analytical hill shading images of this and other DEMs for forested areas which have been produced using the GPM II that the elevations are generally those of a **tree-top surface** rather than of the ground surface itself. The evidence is that road, power-line, and pipe-line cuts are clearly visible as "grooves" in the surface. Thus, the errors referred to above may be correct representations of the surface which was sensed by the GPM II. Whether the pits represent clearings or low areas in the forest cover, or are simply measurement errors, is hardly relevant in the present context: they clearly are not part of the drainage net.

Two approaches can be taken to solve this problem. One approach would begin with the image depicted in Figure 1, and use image processing techniques to detect connected components, and then to eliminate isolated points or connected components below some size threshold. Alternatively, the DEM could be filtered or smoothed to remove or reduce the errors before program HILO is applied. Only the latter approach has been explored thus far.

A simple binomial smoothing operator (Tobler, 1966) was applied once and then a second time to the Keating Summit DEM. Program HILO was then used to estimate the drainage network from these once- and twice-smoothed elevation models; the results are shown in Figures 2 and 3, respectively. The smoothing clearly improves the performance of the algorithm. One pass of the filter eliminates most of the isolated dots, while a second pass removes most of the rest (see Table 1). However, the second smoothing also begins to break up some of the well-defined channels; for the Keating Summit area, a single pass of the smoothing operator seems to produce the best results.

AUTOMATED DETECTION OF DRAINAGE NETWORKS: A HYDROLOGIC APPROACH

Hydrophysically, the drainage network represents those points at which runoff is sufficiently concentrated that fluvial processes dominate over slope processes. If the spatial concentration of surface runoff is simulated, then those points at which this runoff exceeds some threshold can be considered to be the drainage network. This approach has considerable appeal to the geomorphologist, since it is based on physical conditions related to processes. Furthermore, the effects of the chosen threshold on the geometry and topology of channel networks would represent an interesting and as yet unexplored area in network reserach.

Speight (1968) applied this approach manually. First, a square grid of points spaced 100 feet (30.5 m) apart (ground scale) was drawn on a contour map. Next, a slope line perpendicular to the contours was traced downslope from each grid point. Finally, a line segment the same 100 feet (30.5 m) in length was successively centred, parallel to the contours, on each grid point, and the number of slope lines crossing it was determined. If this count exceeded 100 lines, or if the line had a density over a "narrow zone of flow concentration " of more than one slope line per foot of sampling line, the point was declared to lie on a "watercourse" (Speight, 1968, p. 244).

Speight's approach can be automated rather directly for gridded elevation data. Each cell is considered to drain to whichever of its eight neighbours has the steepest downslope toward it. This is not always the lowest neighbour, since the height differences for the four diagonal half-neighbours of a point must be divided by the grid spacing multiplied by the square root of two. Initially, each cell may be considered to produce a unit quantity of runoff; this runoff is then carried downslope in accordance with the drainage directions of the grid cells. Any imported runoff is exported by a cell, together with the locally produced runoff. Then, whenever the runoff in a cell exceeds some threshold, the cell is considered to be part of the drainage network.

Two distinct algorithms for runoff simulation are possible. One requires random access to the elevation data, or at least to the slope directions, and also requires a runoff counter for every cell in the DEM. Each cell is visited, and the runoff unit produced in that cell is followed downslope to the edge of the DEM or to a pit. All cells along the way have one unit of runoff added to the runoff already present. Since runoff from any cell could potentially flow entirely across the data area, the elevations and the runoff counters for the entire model must be available at all times.

A second approach to runoff simulation is local, and each cell is visited only once; however, this algorithm requires that the points be processed in order of decreasing elevation. If cells are processed from highest to lowest, then by the time a cell is processed, all possible inputs to that cell will have been defined (since a cell cannot receive input from one which is lower). In this approach, a sequential file is first written which contains the elevation of each point, together with the cell to which it drains; this file must then be sorted in decreasing order by elevation. A runoff vector with a length equal to the number of elevation points must be initialized to 1, and then the points are processed in order. The runoff present in each cell is simply added to the runoff present in the cell toward which it drains; the runoff need not be carried farther at each step, since all the inputs of any cell are determined before that cell is processed.

The current research was conducted using an Eclipse S/130 minicomputer, and the sorting of a file containing some 160,000 records was considered to be impractical. Therefore, the second approach was adopted. The FORTRAN program DRAIN achieves random access to the elevation data through the use of a Full Morton contiguous data file (Lauzon and others, unpublished manuscript) in conjunction with a virtual array emulator. Briefly, the row and column of a cell are combined to produce a Morton number (Morton, 1966; Lauzon and others, unpublished manuscript); this is then used as an address in the virtual array. Points are brought into core in blocks of 256 elevations. Morton addressing maximizes the probability that the next point required will already be in core, assuming that it is a spatial neighbour. Because of limited core space, the image buffer of a Ramtek color monitor was used to count accummulated runoff. The unit runoff initially placed in each cell was followed downhill to a pit or the map boundary, each pixel along the path was examined, and if its runoff value ("colour") was less than 15, 1 was added to this "colour". Because the image buffer has only four bit-planes, the maximum runoff which could be distinguished was 15; for the present test, then, 15 cells, or 13,500 square metres, was used as the critical drainage area threshold.

Again, pits in the data, which are present due to errors, cause serious problems. Since pits do not have any lower neighbours to which they would drain, runoff accummulation stops at these points, and would begin again from 1 just down-slope. While the runoff of a pit could be exported to its lowest neighbour, this could easily lead to an infinite loop, since that lowest neighbour could simply drain back toward the pit again. Finding the pale, or pour-point, at which water collecting at a pit would flow from the dale surrounding that pit is not a simple task. Once again, smoothing the DEM to remove pits, or at least reduce their number, would appear to be a more practical solution.

EVALUATION

The local concavity approach (program HILO) is much more space- and time-efficient. The method is strictly local, and the single pass through the DEM requires less than 3 minutes on the Eclipse. On the other hand, the runoff simualtion approach (program DRAIN), as currently implemented, is extremely slow, requiring more than an hour of processing time for a single DEM. This is in part because a peripheral device, the Ramtek image buffer, is used to keep track of the runoff, and there is a great deal of costly I/O. Each of the approximately 160,000 cells in a DEM is visited from 1 to 16 times, and at each visit, a "read image", a "move", and a "draw" instruction must be executed. In the present hardware environment, the principal appeal of the DRAIN approach, the possibility of choosing a physically-meaningful threshold for stream sources, is severely restricted; coupled with the extreme slowness of the algorithm, this indicates that at present the HILO approach is to be preferred. However, the DRAIN method still has the potentail to make a greater contribution to our understanding of the realtions among topography, runoff, stream sources, and drainage network structure.

Critical and quantitative comparisons must be made, not only with networks obtained from maps using "traditional" methods, but also with channel networks mapped in the field. The stream patterns identified using either of the algorithms discussed in this paper must also be converted from raster to vector form, in order that link lengths, junction angles, and other properties of interest can be determined. If that result is achieved, data collection should become much less of a limiting factor in drainage network research.

ACKNOWLEDGEMENTS

This reseach was supported by State University of New York Campus Research Development Allocation Grant 150-9762C. Marc Hugus did some initial programming on the DRAIN algorithm, while the filtering routines were programmed by Lloyd Leifer and Peter Liu. All computing was conducted at the Geographic Information Systems Laboratory, SUNY at Buffalo, Duane F. Marble director.

REFERENCES

Allam, M.M., 1978, DTM's application in topographic mapping. Proceedings, Digital Terrain Models (DTM) Symposium, ASP/ACSM, St. Louis, Missouri, May 9-11, 1978, pp. 1-15.

Allder, W.R., Caruso, V.M., Pearsall, R.A., and Troup, M.I., 1982, An overview of digital elevation model production at the United States Geological Survey. Proceedings, Auto Carto 5, (Fifth International Symposium on Computer Assisted Cartography and ISPRS IV Proceedings 1982), pp. 23-32.

Elassal, A.A., 1978, U.S.G.S. digital cartographic file management system. Proceedings, Digital Terrain Models (DTM) Symposium, ASP/ACSM, St. Louis, Missouri, May 9-11, 1978, pp. 16-23.

Jarvis, R.S., 1977, Drainage network analysis. Progress in Physical Geography, v. 1, pp. 271-295.

Kikuchi, L., Guevara, J.A., Mark, D., and Marble, D.F., 1982, Rapid display of digital elevation data in a mini-computer environment. Proceedings, ISPRS Commission IV Symposium 1982, (Fifth International Symposium on Computer Assisted Cartography and ISPRS IV Proceedings 1982), pp. 297-307.

Krumbein, W.C., and Shreve, R.L., 1970, Some statistical properties of dendritic channel networks. Technical Report No. 13, U.S. Office of Naval Research, Contract Nonr-1228(36), Task No. 389-150, Department of Geological Sciences, Northwestern University, and Special Project Report, National Science Foundation Grant GA-1137, Department of Geology, University of California, Los Angeles.

Lauzon, J.P., Mark, D.M., Kikuchi, L., and Guevara, J.A., Two-dimensional run-encoding for quadrtree representation. Sumbitted to: Computer Vision, Graphics, and Image Processing.

Lubowe, J.R., 1964. Steam junction angles in the dendritic drainage pattern. American Journal of Science, v. 262, pp. 325-339.

Mark, D.M., 1983, Relations between field-surveyed channel networks and map-based geomorphometric measures, Inez, Kentucky. Annals, Association of American Geographers, v. 73, no. 3, in press.

Morton, G.M., 1966, A computer oriented geodetic data base, and a new technique in file sequencing. Manuscript, IBM Canada Limited, March 1, 1966.

Peucker, T.K., and Douglas, D.H., 1975, Detection of surface-specific points by local parallel processing of discrete terrain elevation data. Computer Graphics and Image Processing, v. 4, pp. 375-387.

Speight, J.G., 1968, Parametric description of land form. In Stewart, G.A., editor, Land Evaluation, Papers of a CSIRO Symposium in Co-operation with UNESCO, 26-31 August, 1968, pp. 239-250.

Swan, R., Thompson, J., and Daykin, S.E., 1978, Applications of low cost dense digital elevation models. Proceedings, Digital Terrain Models (DTM) Symposium, ASP/ACSM, St. Louis, Missouri, May 9-11, 1978, pp. 141-155.

Tobler, W.R., 1966, Numerical map generalization and notes on the analysis of geographical distributions. Michigan Inter-University Community of Mathematical Geographers (MICMOG), Discussion Paper 8, University of Michigan, January, 1966.

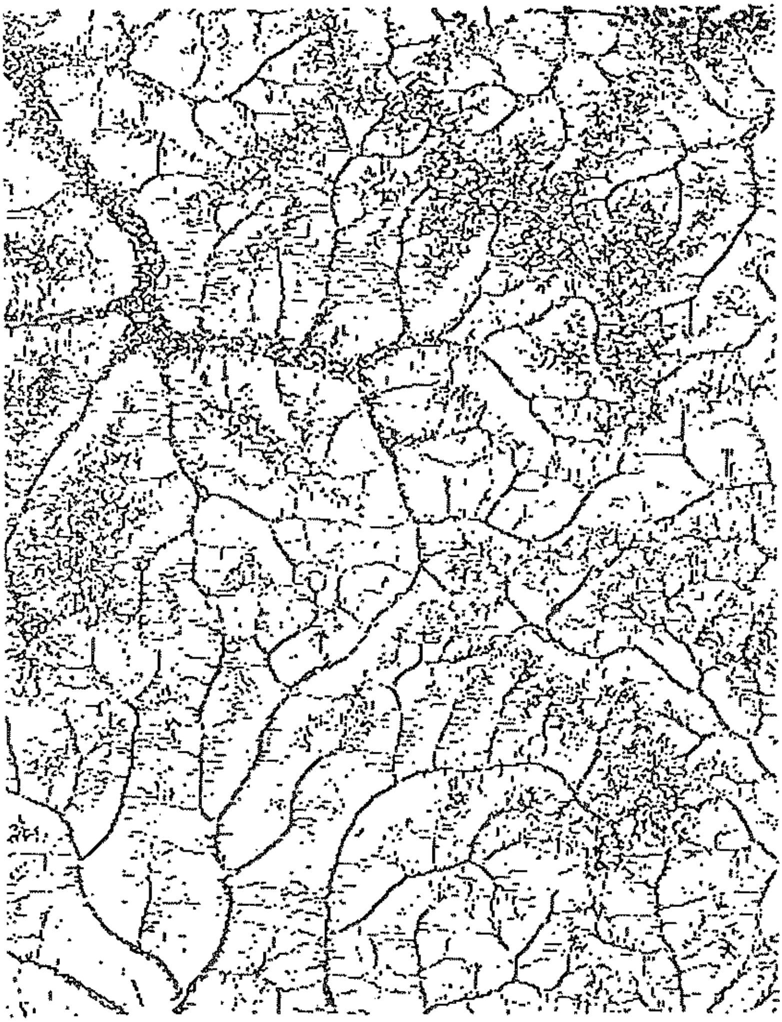


Figure 1. Channel network for the Keating Summit, Pa Quadrangle, based on original (unsmoothed) USGS DEM data (program HILO_).

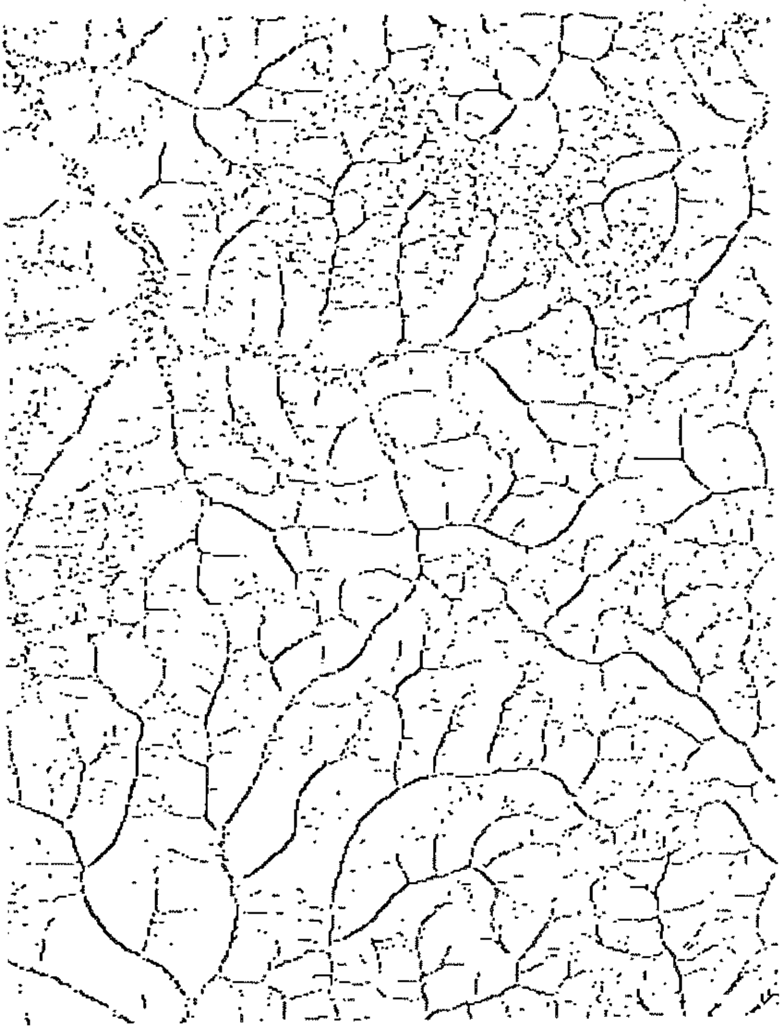


Figure 2: Same as in Figure 1, but DEM data smoothed once before processing by program HILO.

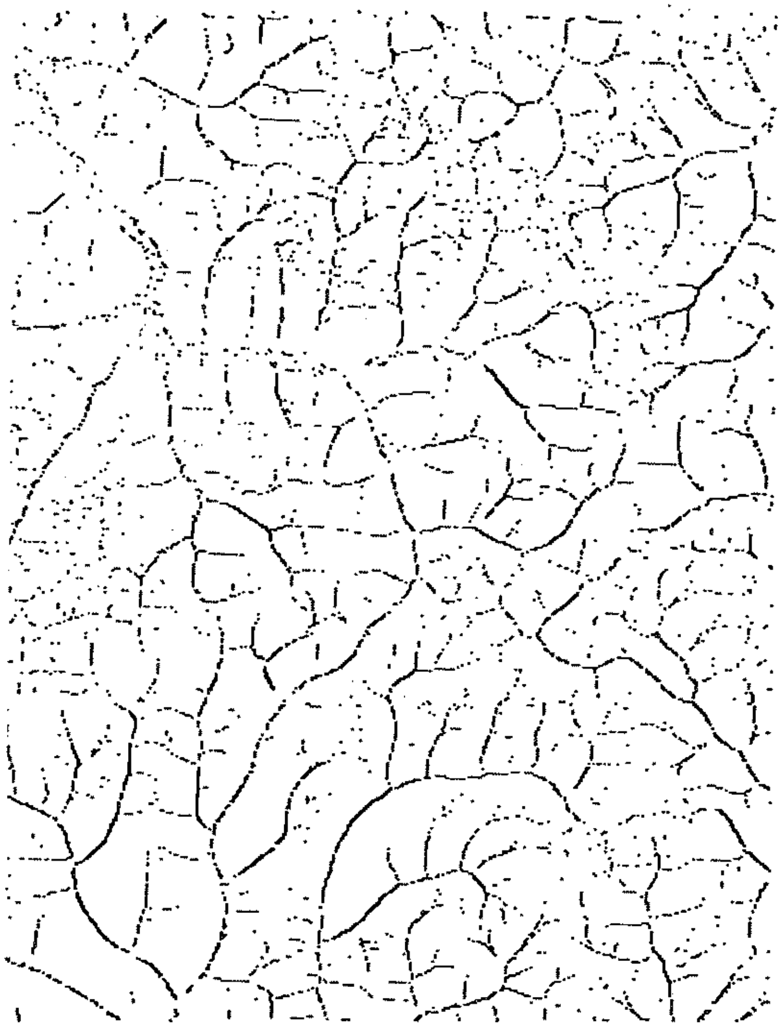


Figure 3: Same as in Figure 1, but DEM data smoothed twice before processing by program HILO.

SHAPE REPRESENTATION BY RECTANGLES PRESERVING FRACTALITY

Pietro Dell'Orco[1]

Marco Ghiron[2]

[1] IBM Scientific Center
v. Giorgione 129 - 00147 Rome Italy

[2] Rome University

ABSTRACT

A new method is proposed for storing drawings defined by polygonals formed by segments of any length by means of a collection of rectangles, each of which circumscribing a different broken line.
Two different approaches are considered, the first one using rectangles with sides parallel to the co-ordinate axes, the second one defining rectangles with any kind of orientation.
The initial data are constituted by the vertices co-ordinates of the broken lines, belonging to the edges of the represented objects.
Initially a thresholding process filters out the segments whose length is smaller than the allowed tolerance, and which can be considered only a noise added to the initial shape. Each polygonal is represented by a series of rectangles, each circumscribing a section of the polygonal.
The length of the least rectangles side must be smaller than the error. This quantity may be either the same for each polygonal or it may be computed each time according to the local features of the polygonals.
To draw the stored shape each rectangle is considered separately and, for each one, a polygonal is found, which presents the highest fractal similarity with the original and the smallest error.
A hierarchy is also proposed to store the same figure with different degrees of resolution, containing a tree in which the rectangles at i-th level are circumscribed at (i-1)-th level.
At the end, methods are suggested for processing intersection, orientation determination, point-in polygon operations.
Comparing this method with those proposed by Loomis [LOOM 65], Burton [BURT 77], [BURT 78], Ballard [BALL 81] we find initial constrains simplified, a new thresholding process, and also a redrawing process which preserves the fractality of the approximated shape.

1. INTRODUCTION

The methods using rectangles to store the original lines are the three ones above mentioned. The most complete is the Ballard's method, which can be considered as an enhancement with respect to the first two methods. From now on only the Ballard's method will be taken into consideration.
The constrains on initial approximation process, imposed in Ballard's method, called regularity conditions, are too restrictive, and so seldom respected. Specifically a regular line is not only required to contain internal broken points or internal intersections, but also its orientation must always be the same (clockwise or not). As the regularity is a local constraint depending on every single segment position and on the defined tolerance, a given polygonal must be analyzed in detail in order to verify whether it satisfies these constraints. In fact it is possible that, accord-

ing to a certain tolerance, a polygonal satisfies the regularity initial constraints, while at a different (and, possibly, higher) approximation level must be subdivided in smaller initial polygonals.
The thresholding algorithms proposed by Ballard are not optimal, because each rectangle does not contain internally the maximum segments number in fact allowed by error constraints. This causes, during update operations, an excessive fragmentation of trees, so one end up with a complex forest as a representation.
The restoration methods are too simple, as it is not convenient to store a line using a (rectangular) surface if, in the restoration process, this is used only to draw a straight line.
The Ballard's method defines also a tree structure storing the same shape with different resolution levels of details.
However, if the approximation has to be improved the polygonals having local concavities should be considered.However, in this method, it is not always possible to define an upper hierarchy level from the lower one; so the approximation process must be resolved by trial and error.

2. INITIAL CONSTRAINTS

The initial shape, that we are analyzing, is formed by a closed or open collection of polygonals composed by a set of segments of arbitrary length.
The only constraints requested are:

(1) Each polygonal must be internally connected.
(2) No polygonal may contain internally points from which more than 2 different segments start.
(3) In every polygonal there are no intersection points.

Should any polygonal not observe these initial constraints this line must be decomposed in a number of polygonals each satisfying the conditions (1), (2), (3).

3. THRESHOLDING (approximation) PROCESS

In this process the irrelevant noisy segments regarding the original polygonal must be filtered out.
Two different thresholding processes are described, the first one filters out the polygonals by using rectangles with sides parallel to the co-ordinate axes (RPAM), while the other one uses rectangles with sides not necessarily parallel to the axes (RASM).
After having divided the initial polygonals into parts, according to the above criteria, each polygonal is separately analyzed and is associated with an error whose value may be the same or it may be computed each time. The thresholding algorithm is computed by finding the rectangles one circumscribing a distinct part of the current polygonal obeying the constraints described in sec. 3.1.
To describe a polygonal using the smallest number of circumscribing rectangles it is better to use a backward algorithm. Each time the current part of the polygonal is included between the current point (Pi) (the starting point coincides in the beginning to the first polygonal vertex), and a polygonal vertex (Pr) which has either the x or y distance from Pi smaller than the error. If the polygonal Pi-Pr does not satisfy the approximation constraints (see 3.1) at every iteration the same line without the last segment is considered.
Therefore applying a forward algorithm it is impossible each time to find the rectangles circumscribing the greatest number allowed of segments (Ballard uses a forward algorithm) as the 4b constraint (sec 3.1) depends on the local properties of the analyzed polygonal. If a point (Pa) lies on a segment whose orientation is different from the contiguous it may be that only in

that point 4b constraint is not verified. So the algorithm stops in Pa instead of going forward analysing other points which satisfy the 4b constraint.

3.1 Thresholding Algorithm

(1) Step: Error E is computed, defining the thresholding level, according to current line.

(2) Step: Point out the initial vertex of the polygon and call it Pi.

(3) Step: Draw the rectangles in the following way:
- (a) The sides are parallel to the co-ordinate axes
- (b) They circumscribe those broken lines whose starting point is Pi.
- (c) They coincide with a growing number of adjoining segments belonging to the initial polygonal.

(4) Step: Choose the rectangle considered in (3) satisfying the following constraints:
- (a) It contains the greatest number of segments
- (b) The extreme points of the contained broken line are on its sides.
- (c) At least one rectangle side must be smaller than the error defined at step (1).

(5) Step: Point out the new initial point Pi by considering the last vertex of the contained line defined at step 4
- (a) If the point Pi is not the last polygonal point we go to step (3)
- (b) Else the thresholding process stops.

If a single segment is greater than the error it is defined by a single circumscribed rectangle.
If a broken line satisfies 4a, 4c but not 4b, and the last point of the broken line coincides with the last polygonal point, it is, all the same, approximated by only one rectangle.
When we approximate a shape by using RASM the best approximation is obtained, as the rectangle circumscribing the polygonal is, at each step, the smallest one. In this way the considered tolerance value can be higher than in RPAM for a given approximation level. Unfortunately the number of data stored for each rectangle are greater and the computing time and the difficulties met increase. In RASM we must modify the third step of the above algorithm as follows:

(3) Step: Consider the following rectangles:
- (a) They circumscribe those broken lines whose initial point is Pi.
- (b) The orientation of a couple of circumscribed rectangles sides must coincide according to the line linking the extreme points of considered broken line.

4. ERROR DEFINITION

If we apply RPAM, the error value is found as follows:
- (a) Calculate the difference between the x and y values belonging to the extreme points of each segment .
- (b) We consider for each segment the smallest difference obtained at step a.
- (c) Compute the arithmetic or weighted average of the values obtained at step b, and this value is the maximum allowed error.

As the error value represents the smallest rectangle side and then the highest allowed distance between the stored and the original polygonal, it is many times greater than the side length of the lattice. In particular cases a formula proposed by Pavlidis [PAVL 82], can be applied conveniently also to compute the tolerance value in RASM:

$$e(max)=(D/2)[(S/D)^2-1.1]^{\frac{1}{2}}$$

(D=distance between extreme points, S=curve length). Taking the length of the rectangle side between 0 and e(max)/2 this value will be much smaller as better or more accurate is the approximation.

5. STORING PROCESS

After the thresholding process we can choose between two possibilities:

(i) If we want to have only polygonals as output from the thresholding process, only a redrawing process (later described) is applied.

(ii) If we want to store the rectangle structure, and possibly to redraw the polygonals afterwards, also a storing process must be considered.

Using RPAM the following data are stored:

(1) Polygonal identifier the rectangle belongs to.

(2) The rectangle length and height.

(3) Co-ordinates of the low-left rectangle corner.

(4) Co-ordinates of the initial and final point of the contained broken line defined using the curvilinear co-ordinates on the rectangles sides starting from the low-left rectangle vertex.

Instead using RASM the following data are stored:

(1) Polygonal identifier the rectangle belongs to.

(2) Extreme points co-ordinates of the broken line.

(3) Three more data describing the distance between the extreme points of the broken line and the rectangle sides. These values can be negative, positive or null depending on the cases we can meet.

These data describe biunivocally every polygonal part and they allow to redraw the thresholded shape.

6. REDRAWING PROCESS

The redrawing process transforms the stored data describing the circumscribed rectangle into a shape composed by polygonals lines. The stored rectangles describe the area filled by the initial broken line, and therefore a polygonal can be found, which presents the highest fractal similarity with the original and the smallest error.

The error defined by the distance between the original polygonal and the redrawn polygonal will be certainly less than E/2, or even negligible.

This redrawing algorithm connects the initial and the final points belonging to each rectangle through a broken line whose vertices belong to the rectangle sides.

Each polygonal is separately drawn, and its data are initially gathered and ordered in an order which is the same as the one obtained from the thresholding process.

6.1 Redrawing Algorithm

First, for each rectangle, compute the number of sides touched by Pi and Pe (extremes of each thresholded polygonal), then apply, for each rectangle one of these steps:

Cases:

(1) 4 sides: The vertices of the redrawn broken line are Pi and Pe (see figure 1c: r1,r2).

(2) 3 sides: Either Pi or Pe coincides with one of the rectangle vertices, while the other one lies on the opposite side. The broken line vertices are: Pi, the middle point of the only rectangle side which neither Pi or Pe belong to, and Pe (fig 1c: r3) .

(3) 2 sides: (i) the sides which Pi and Pe belong to are adjoining, the vertices are: Pi, the middle points of the sides which neither Pi or Pe belong to, and Pe, in the order Pi to Pe.

(ii) The Pi and Pe sides are opposite: The broken line vertices are: Pi, Pr, Pl, Pe. Pr belongs to that side perpendicular to Pi side, whose distance from Pi is the smallest. The Pr distance from Pi side is equal to the length of the side it belongs, multiplied by 1/4 (call this distance DPr). Pl belongs to

the side opposite to Pr side and its distance from Pi side is DPr multiplied by 3 (fig. 1c: r4). These parameters are useful for preserving fractality as described in the following paragraph.
(4) 1 side: The broken line vertices are: Pi, and the middle points of the sides which Pi does not belong to, and Pe.
When the final point is inside the rectangle the broken line points are: Pi, the middle points of the other rectangle sides, and the rectangle center.
When we adopt RASM we can entirely apply the described redrawing process, even though the stored data are defined in a different way.

6.2 Some Considerations on the Redrawing Process

In order to ensure continuity between the broken lines redrawn from the stored rectangles belonging to the same initial polygonal, each broken line must pass always through its extremes Pi and Pe previously stored.
In order to find the remaining internal points in the circumscribed rectangles we follow these criteria:
(1) If one rectangle side does not contains Pi, or Pe, it must contain at least one intermediate point of the approximating polygonal, as the stored rectangle is the smallest.
(2) The redrawn line approximates the more the original one the more is his fractality; nevertheless, after a few iterations, a similarity near to the asyntoptic one is reached. Therefore, it is convenient to keep this fractality not so high. (Reasonable values for fractality are between .2 and .5, see [MAND 77], [FOUR 82]).
A fractal polygonal is a line which preserves a peculiarity called fractal dimension D, which defines the irregularity of the line itself [MAND 77].
Generally a recursive algorithm is used to draw a fractal line, at each iteration a new point is found whose distances from the two previous computed points (the initial points are the extremes of the polygonals) and whose variation, from the straight line connecting these two points, depends on a factor d.
In the redrawing process the factor D is not exactly maintained but it is approximated to preserve only the principal features of the stored shape.
For instance in the step 3b, in the preceding algorithm, the point Pr is chosen on the side nearest to Pi, as is that variation the most probable, and the fractal Brownian motion (on which the fractality is based) is generated using a gaussian process with 0 average and unit variance multiplied by a scale factor.
Considering fractality constraints is necessary to obtain a good approximation, a clear example of the contrary is figure (4). We see in fact that the redrawn line does not follow in a right way the initial polygonal, while the other line (F), obtained using fractality constraints, is a better approximation.
The polygonal to be redrawn is the one which has a uniform distribution of the fractality in the rectangle, as it is the most probable, and this fact can be seen in many examples in the cartography.

7. HIERARCHICAL STRUCTURE

We propose a hierarchical structure defined on the above methods, allowing to describe the same shape using different resolution levels.
We build a binary tree, whose first level (leaves) describes the circumscribed rectangles associated to the initial polygonals with a defined error value.
At upper tree levels we find rectangles circumscribing the contiguous rectangles belonging to the immediately lower level.
A binary tree structure is being chosen as it has many advantages during approximation and updating process.
The binary tree is composed in this way:

(1) Each no terminal node has two sons, if the nodes number at i-th level is odd, the last node will be considered at (i+1)-th level.
(2) The terminal nodes circumscribe the polygonals by using the previously defined methods.
(3) The root node defines the rectangles circumscribing the entire polygonal. But to be sure to draw always a circumscribed rectangle from the lower level, we must slightly modify the constraints of the previous representation. In fact the extremes of the broken line may also be inside the circumscribing rectangle. Not to increase the number of stored data, we only consider new possibilities in RPAM, if either Pi, or Pe is inside the rectangle we store its position in the following way:
Consider the following rectangles:

(i) The first one (i=1) coincides with the rectangle circumscribing the polygonal

(ii) The other ones are parallèl to co-ordinate axes and the (i+1)-th rectangle is included in the i-th one, and its distance from the previous one is equal to unity.

The point P to be stored is surely on one of those rectangles and its curvilinear co-ordinates are found by applying this formula:

$$D=(((i-1)*l)-((i-1)*8)+m$$

where:

i is the identifier of the i-th rectangle
l is the length of the i=1 rectangle
m is the curvilinear distance from the lowest rectangle vertex on the i-th rectangle which P belongs to.

If both points Pi and Pe are enclosed in the rectangle we must also consider the polygonal orientation (see orientation algorithm) between these points, which can be stored using a sign (+,-) in the polygonal identifier.
This case possibly holds starting from the third and higher tree levels, and it is not considered in Ballard's method. In Ballard's method the regularity constraints ensure that Pi and Pe are always on the sides, but it is difficult to maintain always these conditions. For instance in figure 5 the rectangle circumscribing the entire polygonal has the extreme polygonal point on its sides, instead, considering the smaller rectangle at a lower tree level circumscribing the broken line CB the extreme point B lies internally. So using Ballard's method the initial polygonal must be divided in at least two parts. Then the representation process must be again executed completely. Instead using RASM or RPAM methods the representation process may be completely applied in order to find for each initial polygonal only a tree and not a forest of trees.
Note that when both Pi and Pe are inside the rectangle, the redrawn polygonal has the the same stored orientation and a vertex on each rectangle side.

8. RESULTS

Using the RPAM or RASM we can reduce considerably the initial stored data quantity. In figure 1a the original polygonal, formed by 1126 vertices, is drawn; it is stored in a array whose dimensions are (1126,5). After the representation process (RPAM) fig. 1b (error 1000) we have an array whose elements are (37*7), which can be redrawn as in fig. 1c. Comparing the redrawn (R) polygonal to the original (O) polygonal we see that there is an enormous advantage in storage reduction and only a small error between the two polygonals. If we choose an error equal to 500 we obtain an array of (77*7) elements (figs. 2b, 2c, and the O one in fig 2d) instead the one having (1120*5) elements (fig 2a, the F one in fig. 2d) and a better approximation. Using the RASM we see in fig 3a the original polygonal formed by an array, whose dimensions are (1120*5) we obtain a good approximation using an error equal to 1500 and having in this way an array (60,8).(figs. 3b, 3c, comparison in fig. 3d).

9. INTERSECTION ALGORITHM

This algorithm finds the intersections between a segment S1 and one stored polygonal. The middle part of this algorithm is similar to the one described by [LOOM 65], the initial and final steps are totally different because only in this way all the intersection possibilities can be correctly examined.
This algorithm is formed by six steps, during which the number of intersections are continuously reduced, until the sure intersection points with segment S1 are found.
(1) The circumscribing rectangle S1 with sides parallel to the axes (R1) is determined.
(2) Find the set Ir containing the rectangles whose intersection is sure:
To test that property we find the following polygon A1:
  (i) Consider the strip ST whose sides are coincident with the sides of the current stored rectangle (M1), and parallel to the x-axes.
  (ii) A1 is obtained considering the region (ST-M1)
If the extremes of M1 are in A1 and the first one lies in the region opposite to the one of second vertex with respect to M1, then a sure intersection is found. Repeat step (2) considering at (i) the strip sides parallel to y-axes. Each rectangle belonging to Ir and whose test is positive, is stored in Isure set.
(3) Now find only possible intersections: compare rectangle R1 with the ones in Ir (and vice versa). There is a possible intersection if at least one of the rectangle vertices is included in a Ir rectangle; those rectangles satisfying this constraint are stored to be analyzed in the following step.
(4) The polygonals circumscribed by the stored rectangles are redrawn. Test if a vertex of every polygonal is inside the rectangle R1, and then check for the inclusion of S1 vertices in the rectangles circumscribing the redrawn segments. The segments whose result is positive are collected in Ipossible set.
(5) The intersections between S1 and Isure or Ipossible are analytically computed, paying attention when the segments are coincident or parallel.
(6) The sure intersections are those found at step 5 from Isure set or from Ipossible set whose intersection point is inside R1.

10. LINE ORIENTATION

This algorithm checks for the orientation of a polygonal line, or the direction of a closed line connecting two points, comparing only a little part of the polygonal segments. We only analyze some important features of some of the points, which lie on one side of the rectangle circumscribing the closed curve.
  (1) Consider the stored circumscribing rectangles and we find the ones (Ir set) lying on the side of the rectangle circumscribing the closed polygonal.
  (2) Consider the Ir rectangle whose y-value is the highest
    (a) Follow forwards the chain from that rectangle until find another Ir rectangle
    (b) Repeat step 2a going backwards
  If the x variation from rectangle 2a to 2b is positive then the polygonal orientation is clockwise, else it is anti-clockwise.
  (3) Comparing the direction of a line Pa-Pb with the polygonal orientation, we can also investigate the direction between two points.

11. POINT IN-POLYGON OPERATION

We suppose that the investigated closed line is simple: it has no internal nodes or break points, and we consider to have as initial data the ones stored by RPAM. Making only some slight variations we may also apply this algorithm to every kind of initial data.

(1) Find the rectangle circumscribing the closed polygon, on the base of the vertices of the stored rectangles. Then test if the current point Pa is inside this rectangle:

(i) If the test is positive we go to step 2

(ii) else the point is certainly outside the polygon and the algorithm stops.

(2) Find the stored rectangles whose y co-ordinates values include the Pa y-value (I1 set).

(i) Pnt into I2 set the I1 rectangles which include Pa. Now find the part of I2 rectangles inside the polygon. Applying the orientation algorithm we find the orientation of the polygon. The part of the of the rectangle inside the polygon is the one found by drawing the closed figure whose sides are composed by the polygonal redrawn by the stored rectangle from Pi to Pe, and the rectangle sides followed from Pe to Pi having the same polygon orientation. Now we can test if Pa is inside this simple figure (Note that the maximum number of sides is six).

(ii) Pa conld be inside the polygon and outside the stored rectangles.

Consider the two I1 rectangles, whose distances from Pa are the smallest from the left or the right side (R1, R2), (We could even consider the entire I1-I2 set).

Draw from R1 or R2 the polygon whose borders are:

(i) The circumscribed rectangle side parallel to the y-axes having the same polygonal orientation between point Pi and Pe

(ii) The half-lines parallel to the x-axes and whose initial points are (min(Xi,Xe),min(Yi,Ye)) or (max(Xi,Xe),max(Yi,Ye)) where Pi=(Xe,Ye) Pi=(Xi,Yi) if the internal polygon part is on the right or the left side of the rectangle. (orientation algorithm).

(3) If Pa is inside two nearby rectangles previously found it is certainly inside the polygon itself.

12. CONCLUSIONS

The RPAM is a simple and efficient method because the main operations to be computed are distance operations which may be resolved by simple differences between x or y co-ordinates of the polygonal vertices. Due to this fact the polygonal interval, in which the circumscribing rectangle may be defined, is immediately found.

Even if the RASM defines a better approximation level it is easier to apply RPAM.

Due to a good redrawn process, which also redraws polygonals preserving fractality and not only straight lines, it is possible to get a very good storage data compression (.8 or even more) (see sec.8).

According the initial constraints it is always possible to define one level of the tree from the lower, and the data stored are (2**n)+(2**(n-1))..(2) (where ** is the exponential operator) and where the data (m) stored at the first level are (m) (2**n).

When a part of a stored polygonal is updated it is necessary to consider, in the thresholding process, only the new polygonal part. As the number of data about the entire polygonal does not change by considering a new thresholding process on the entire polygonal or only on the new part. In fact the fragmentation process is absent: as soon as, during the thresholding process, a

previously stored rectangle, whose dimensions are equal tolerance value, is met, the fragmentation process stops. The rectangles obtained by considering those two storing possibilities are only shifted with respect one another.
The geometric algorithm proposed can be resolved by computing only simple arithmetic and logic operations between the stored data. Generally the stored data which must be analysed are less than .25 and, for a smooth shape, less than .02 .

REFERENCES

[BALL 81] H. Ballard
Strip tree: a hierarchical representation for curves
Communications of ACM
Vol.24, N.5, may 81

[BURT 77] W. Burton
Representation of many-sided polygons and polygonal lines for rapid processing
Communications of ACM
Vol.20, N.3, march 1977

[BURT 78] W. Burton
Efficient retrieval of geographical information on the basis of location
Communication of ACM

[FOUR 82] A. Fournier, D. Fussell, L. Carpenter
Computer rendering of stocastic models
Communications of ACM
Vol.25, N.6, June 1982

[LOOM 65] R.G. Loomis
Boundary networks
Communications of ACM
Vol.8, N.1, January 1965

[MAND 77] B.B. Mandelbrot
Fractals form, chance, and dimension
W.H. Freeman and Company Editor

[PAVL 82] T. Pavlidis
Curve fitting as a pattern recognition problem
Sixth international conference on Pattern Recognition Munich, Germany, October 19-22 1982

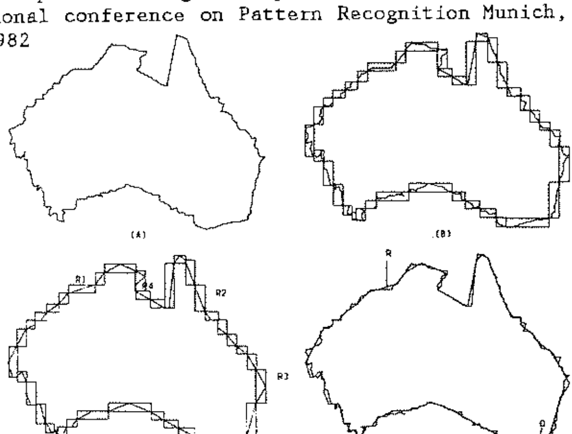


FIGURE 1

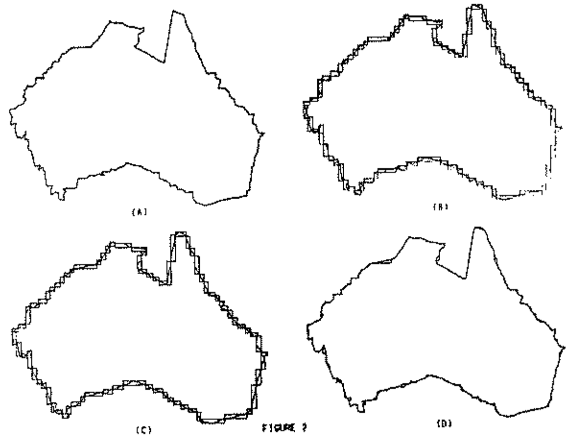


FIGURE 2

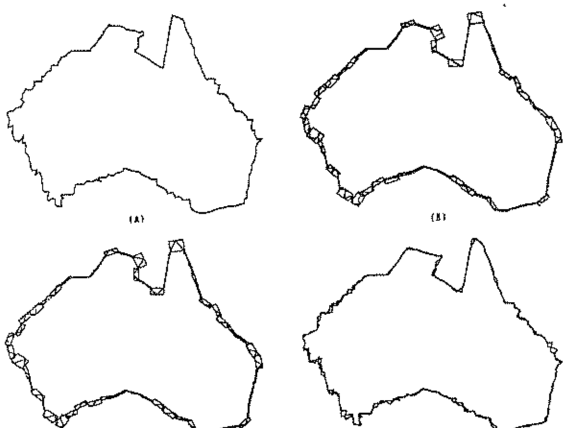


FIGURE 3

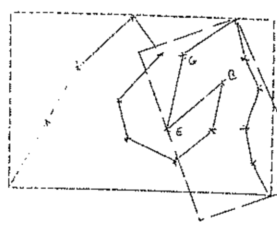


FIGURE 5

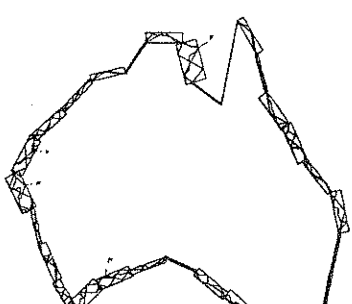


FIGURE 4

FRACTAL ENHANCEMENT FOR THEMATIC DISPLAY OF TOPOLOGICALLY STORED DATA

Marc P. Armstrong and Lewis D. Hopkins
Department of Landscape Architecture
1301 West Gregory Drive
University of Illinois
Urbana, Illinois 61801

ABSTRACT

Most geographic information systems reference data using coarse grid cells or explicitly encode areas using high resolution coordinates. In certain circumstances it is more efficient to reference data solely in terms of the topological relationships among data entities. This approach, however, does not provide for adequate map display of data. If at least one of the types of entities in a database is geocoded, even if very coarsely, then it may be practical to display other data with respect to these coordinates. An approach has been implemented in which chains of centroids of United States Public Land Survey sections are enhanced through fractalization to provide thematic maps of streams.

INTRODUCTION

Most geographic information systems reference data using grid cells, or explicitly encode areas using high resolution coordinates. Data capture, storage and retrieval for such systems can be expensive if topological relationships must be determined at high resolution. For example, determination from coordinates of left versus right bank of a stream is difficult without very high resolution data. Rather than use spatial referencing procedures, in certain circumstances, it is more efficient to reference data solely in terms of topological relationships among data entities. A major disadvantage of this approach, however, is that it does not provide for adequate map display of data, even for thematic mapping purposes. If at least one of the types of entities in a database is geocoded (e.g. coordinates of its centroid), then it may be practical to display other data with respect to these geocodes, even if they form only a coarse net. This paper describes such an approach in which chains of United States Public Land Survey (USPLS) sections that are contained in a streams resource information system, are enhanced through fractalization to provide thematic maps of streams.

ILLINOIS STREAMS INFORMATION SYSTEM

The mapping system described here is but one component of the Illinois Streams Information System (ISIS). ISIS is a computerized data storage and retrieval system for the 2,000 Illinois streams with watersheds of 10 square miles or larger. The database contains a wide variety of data items including locational, biological, chemical, physical, cultural and recreational variables (Illinois Streams Information System Staff, 1982). An important characteristic of ISIS is the use of existing database management software to implement efficiently a data structure. The ISIS database uses a network database model, with a number of inverted lists to improve data access and user friendliness.

Topological Referencing in ISIS

Retrieval efficiency in ISIS is achieved by storing data in terms of stream network relationships through the use of a hierarchical stream numbering system. In the hierarchy, each stream is described by a sequence of numbers, such as 10-30-20, where 10 indicates the final outflow stream (e.g., the Mississippi); 30 indicates the third tributary of stream 10; and 20 indicates the second tributary of stream 10-30. The zeros are used as place holders for future additions to the hierarchy. An entire watershed can be identified from these numbers simply by retrieving information for all streams numbered for example, from 10-30-00 through 10-30-99.

Within streams, data are locationally referenced in ISIS by the distance measured along each stream from its mouth - the River Mile Index (RMI). RMI data were obtained from the United States Geological Survey (Healy, 1979), for all stream intersections and major stream related cultural features (e.g. bridges). RMI's for intermediate points are determined by measuring relative to features with given RMI's, thereby avoiding cumulative measurement error. The RMI data allow for the computation of stream network distances or frontages along streams.

The use of both stream numbers and RMI's, which are the primary locational keys in the database, provides for the determination of topological relationships among stream network elements. In addition, data are referenced by right bank, center, or left bank. Although this data structure provides all the topological information necessary for analysis, it precludes locating an object on a traditional map using spatial coordinates. A link that relates the topologically referenced data to a coordinate system is required.

Spatial Referencing in ISIS

ISIS includes such a link to coordinate information. USPLS section, township, and range are encoded for all sections into which a stream flows. These section data are included in order to provide inverted lists for data entry and for data retrieval. The RMI at the entry to each section is also recorded. This link makes it possible to obtain an ordered set of sections through which a stream flows. It provides a good, albeit low resolution, means for spatially locating data because the USPLS system is a mostly systematic partitioning of geographic space. It is generally a grid with adjustments every 24 miles to compensate for the earth's curvature. In Illinois these section data can be processed by the Illinois State Geological Survey's ILLIMAP (Swann et al., 1970) system, which returns the centroid coordinates (in state plane coordinate feet) of each section. This capability can be used to obtain very low resolution, ordered coordinate chains, which provide the skeleton for the ISIS mapping system.

MAPPING PROCEDURES

The low resolution coordinate chains when plotted in their "raw" form, do not resemble a stream network. As shown in Figure 1, a plot of these coordinate

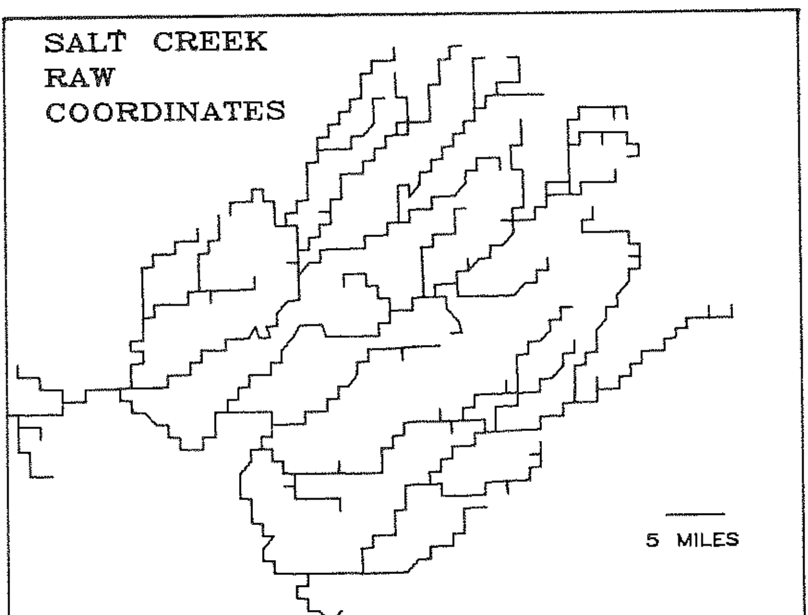


Figure 1. Raw plot of section centroids.

chains is not adequate for thematic mapping because the chains do not retain the stream network topology and they are too rectilinear to comprehend as a stream pattern. In order to make these chains sufficiently resemble streams for thematic maping, they are enhanced using smoothing and fractalizing algorithms. Before they can be enhanced, however, the data must first be processed to delete incorrect stream forms.

Data Cleaning Procedures

Stream widths and meander frequencies are normally much smaller and higher, respectively, than can be effectively rendered by a one square mile grid. For example, when a stream meanders from one section into another and then back into the original section (Figure 2a), the chain of section centroids yields a retraced line segment. In the context of the entire stream, this retracing appears as a spike (Figure 2b).

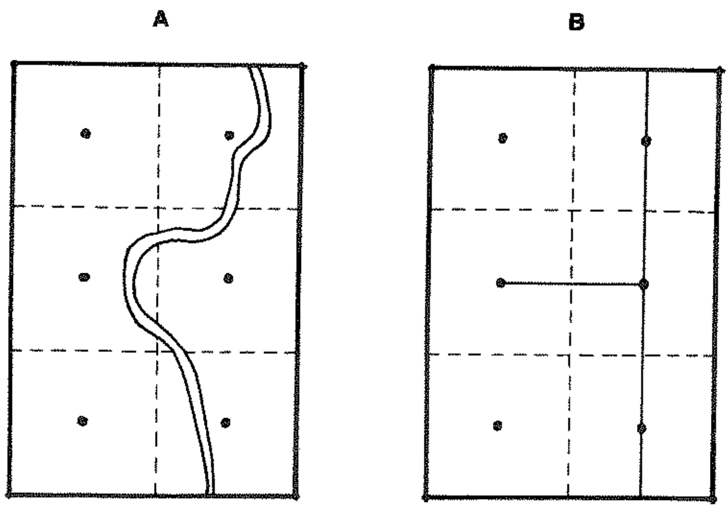


Figure 2

Figure 2. Retraced line segment.

Another problem occurs when two streams enter the same section but do not actually join. Usually they join in a section farther downstream (Figure 3a). This situation yields a cycle in which the streams appear to flow in a closed loop (Figure 3b).

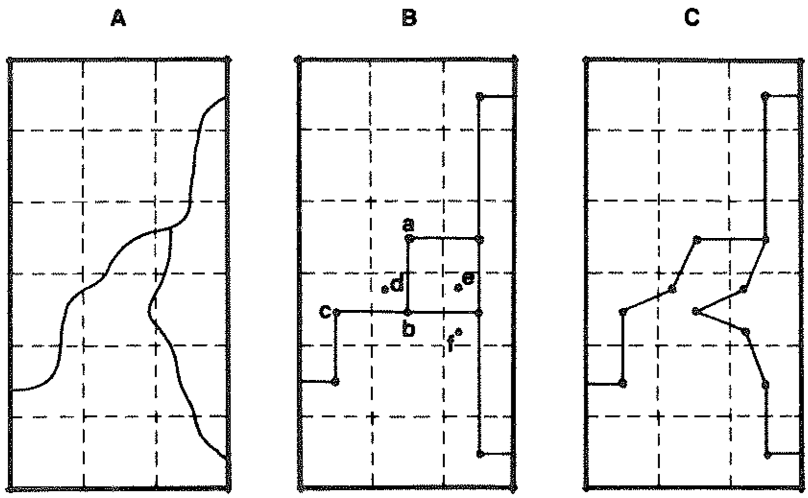


Figure 3

Figure 3. Processing of duplicate points.

These problems were dealt with by shifting the duplicated centroids in the following way. Duplicate points are identified through visual inspection of a plot of the centroid chains and are assigned a numerical flag. When this flag is encountered in the line enhancement program, the upstream and downstream neighbors of the point are identified. As shown in Figure 3b, this three point chain is processed so that the flagged point forms a vertex of the triangle, ABC. A new vertex, D, is created by moving the vertex, B, a proportion of the distance to the base of the triangle along the perpendicular bisector of the angle at the vertex. Thus, the line segments are peeled apart (Figure 3c.) If duplication occurs for a series of contiguous points in a chain, this procedure is applied repeatedly until the chains are entirely peeled apart. The same procedure also works well for altering the aforementioned spikes (Figure 3c). In that case the duplicate vertices, E and F, are peeled apart at the base of the spike.

With these corrections, the steam networks as represented by chains of section centroids, are topologically correct. They are, however, still too rectilinear to read as streams.

Line Enhancement Procedures

A fractal (Mandelbrot, 1977) line enhancement technique is used to introduce small scale features into the cleaned, low resolution line segments. The enhancements are not random forms, but rather are caricatures and recursions of forms found in the line given as input to the fractal algorithm (Dutton 1981: 25). Fractalization of a line is a deterministic procedure.

In this application the input chains, even after being cleaned, do not possess appropriate forms for enhancement by fractalization because they are rectilinear. They are not stream forms generalized to a small scale. For that reason, the usual line enhancement procedure must be altered. In the revised procedure, the chains are smoothed (using splines) before fractalizing as well as the usual smoothing after fractalizing. The resultant higher resolution forms introduced by the fractal algorithm thus gain a stream like character from the smoothed chain of section centroids. This smoothed chain is more reminiscent of small scale stream lines than is the rectilinear chain of centroids.

Although sinuous forms can be introduced readily into a line, the enhanced line should 1) meet expectations as to form and should 2) retain the scaled length of the stream as specified in the database in order to map thematic data by RMI.

The prescribed length can be achieved by fractalizing or smoothing iteratively and terminating when the enhanced line is within a certain tolerance of the actual scaled distance. Although this process is easily implemented, one smooth, fractalize, smooth sequence has yielded lengths generally within acceptable tolerances for present purposes.

Because high resolution features are introduced into the line segments from low resolution data, the resultant lines exhibit self-similarity at different resolutions. If regular or repetitive forms are present in the input chains, these regular forms are introduced into the output chains. The smoothed centroid chains are more regular in form than are actual small scale stream lines. Some unnatural repetitiveness and regularity therefore occurs in the output (see Figure 4). This regularity might be distracting in a thematic map. It could be reduced by using a stochastic operator in the fractal procedure.

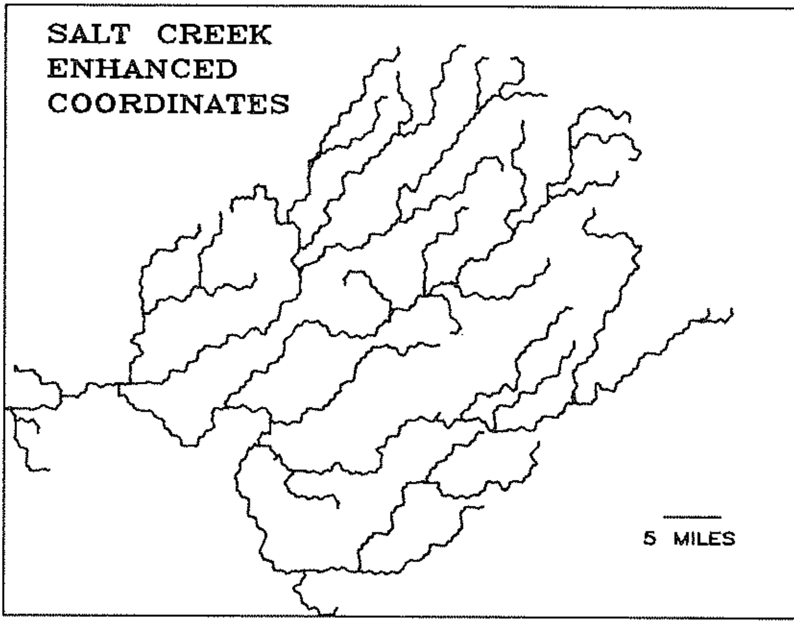


Figure 4. Enhanced section chains.

LINE ENHANCEMENT RESULTS

Figure 1 is a plot of the centroid chains for the Salt Creek River Basin in their raw, rectilinear form. Notice the numerous closed cycles, spikes, and angular features. Figure 4 shows a map of the same streams after they have been cleaned of low resolution artifacts and then enhanced using the procedure described above. The resulting output has a more streamlike character than the input chains, and is a reasonable base upon which to display thematic information.

The parameters used in smoothing and fractalization were chosen based on our experience to mimic a map at a scale of approximately 1:250,000, the scale used for thematic maps for the ISIS project. The figures shown here are greatly reduced so the stream lines are actually fractalized too much for the scale at which they are presented in Figure 4. Plots at 1:250,000 do resemble the stream lines on USGS 1:250,000 maps. Gardiner (1982) and Klein (1982) discuss the effects of scale on sinuosity in stream networks.

THEMATIC DISPLAY OF DATA

After the rectilinear chains have been enhanced, they can be used to plot streams information that is contained in the database. There are two additional steps in creating these thematic maps: 1) scaling the features to be plotted with respect to the enhanced line and 2) generating parallel lines to represent bankside features.

Scaling of Features

Thematic data are stored in the database using either a single RMI for point data, or the beginning and ending RMI for line data. The bank position of each data item is also stored in the database. For plotting, an RMI must be matched to an appropriate coordinate in the enhanced line. The length of the stream is contained in the database and the enhanced line is fractalized and smoothed to be a similar length. For the following procedure streams are broken into segments derived from confluences to avoid accumulating error. Pointers to the two nearest coordinate pairs are calculated by determining the proportion of the relative position of the feature to the length of the segment. From these two coordinates a new coordinate is calculated from the exact proportion of the relative position of the feature to the length of the segments.

Display of Bank Data

Bankside data are plotted by processing the enhanced centerline data to produce new line segments - surrogate banks - which mimic the input line. These surrogates, however, have local lengths different from the length of the stream centerline, and normally are more generalized. Again to avoid accumulating error, the following procedure is applied to each stream segment. The actual RMIs for the bankside data are contained in the database. The length of each displayed bankside line is calculated and the actual RMI scaled to the proportional distance on the surrogate bank. Figure 5 shows a thematic map of bank data.

A more difficult problem occurs at the confluence of two streams, when linear data plotted as banksides

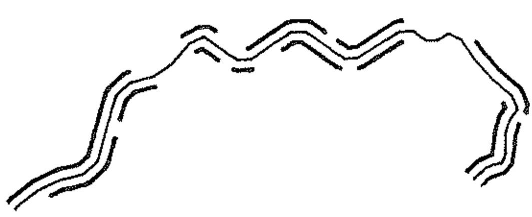


Figure 5. Bankside Vegetation

can actually cross. There are at least four ways to deal with this problem. The first is to let the lines cross, being an artifact of line width and scale. A second is to establish a hierarchical primacy. The bankside data of the highest order stream takes precedence, for example, over intersecting streams. This procedure creates T intersections instead of crosses and implies that the features occurring along high order streams are of greater importance than features along low order streams. A third, and better method is to offset slightly the bankside features at stream confluences. At each stream confluence, bankside plotting begins an increment equal to the distance from stream centerline to a parallel bankside line upstream from the confluence. It ends the same distance downstream from the next confluence. This offset is small in terms of ground distance at the chosen plotting scale. A fourth procedure involves the analytical computation of any bankside feature intersection points. Once these points are calculated, crossover links may then be deleted if they are present. This procedure is less attractive than the previous one, because it entails more processsing (Yoeli, 1982), and does not yield a substantially improved result.

CONCLUSIONS

In developing the Illinois Streams Information System encoding high resolution stream centerline chains was rejected because of high initial costs of data capture, verification, and of software to support a high resolution geographic information system. Others have discussed the difficulties and expense of obtaining clean digitized lines (see for example, Dutton, 1978: 81). Because USPLS section information is an important data item in ISIS and a convenient method exists for converting the section centroids into coordinates, an approach has been implemented that transforms the low resolution chains through smoothing and fractalization into chains that resemble streams sufficiently for thematic mapping. The approach has been shown to provide useful thematic maps at a scale and precision appropriate for ISIS.

A thorough evaluation of the approach will require comparison of costs for data capture, maintenance, storage, retrieval, and display with costs for alternative approaches. Given the need for USPLS section data in the database and the availability of the ILLIMAP system, the approach of encoding and storing data topologically and providing map display through enhancement of low resolution geocoded data entities appears to be a viable alternative to the high, up front costs of a high resolution geographic information system. Some states that have extensive, coarse grid cell databases may find the approach useful in providing display of topologically encoded and stored stream data.

ACKNOWLEDGMENTS

The work reported in this paper is being performed in part under contract with the Illinois Department of Conservation and funded in part by the U.S. Department of the Interior. We wish to thank Geoffrey Dutton of Harvard University for providing the original fractal and smoothing subroutines used in developing our system. Parallel line subroutines were implemented from code written by David Douglas. David Head assisted in the development of the line cleaning procedures.

REFERENCES

Dutton, G. H. 1978, ed., Proceedings of the First International Advanced Study Symposium on Topological Data Structures for Geographic Information Systems, Volume 1 - Summary of Proceedings, Laboratory for Computer Graphics and Spatial Analysis, Harvard University, Cambridge, Massachusetts.

Dutton, G. H. 1981a, Fractal Enhancement of Cartographic Line Detail: The American Cartographer, Vol. 8, No. 1, p. 23-40.

Dutton, G. H. 1981b, Personal Communication.

Gardiner, V. 1982, Stream Networks and Digital Cartography: Cartographica, Vol. 19, No. 2, pp. 38-44.

Healy, R. W. 1979, River Mileages and Drainage Areas for Illinois Streams - Volume 2; Illinois River Basin, U.S. Geological Survey Water Resources Investigation 79-111, U.S. Geological Survey, Champaign, Illinois.

Illinois Streams Information System Staff 1982, Project Description, Department of Landscape Architecture, University of Illinois, Urbana, Illinois.

Klein, P.M. 1982, Cartographic Databases and River Networks: Cartographica, Vol. 19, No. 2, pp. 45-52.

Mandelbrot, B. B. 1977, Fractals: Form, Chance, and Dimension, W. H. Freeman, San Francisco.

Swann, D. H., P. B. DuMontelle, R. F. Mast, and L. H. Van Dyke 1970, ILLIMAP - A Computer-Based Mapping System for Illinois, Circular 451, Illinois State Geological Survey, Urbana, Illinois.

Yoeli, P. 1982, Cartographic Drawing with Computers: Computer Applications, Vol. 8, pp. 1-137.

MEASURING THE FRACTAL DIMENSIONS OF SURFACES

Mark C. Shelberg
Cartographer, Techniques Office
Aerospace Cartography Department
Defense Mapping Agency Aerospace Center
St. Louis AFS, Missouri 63118

Nina Lam
Assistant Professor
Department of Geography
Ohio State University
Columbus, Ohio 43210

Harold Moellering
Associate Professor
Department of Geography
Ohio State University
Columbus, Ohio 43210

ABSTRACT

The fractal dimension of a surface is a measure of its geometric complexity and can take on any non-integer value between 2 and 3. Normally, the topological dimension of surfaces is 2; however, their fractal dimensions increase with greater amounts of complexity or roughness. For example, a fractal dimension of 2.3 is found to be a common value in describing the relief on the earth.

This paper discusses and presents examples of an algorithm designed to measure the fracticality of surfaces. The algorithm was developed at The Ohio State University and is shown to be reliable and robust. It is placed in an interactive setting and is based on the premise that the complexity of isarithm lines may be used to approximate the complexity of a surface. The algorithm operates with the following scenario: Starting with a matrix of Z-heights, an isarithm interval is selected and isarithm lines are constructed on the surface. A fractal dimension is computed for each isarithm line by calculating tbeir lengths over a number of sampling intervals. The surface's fractal dimension is the result of averaging the fractal dimensions of all the isarithm lines and adding 1.

Potential applications for this technique include a new means for data compression, a quantitative measure of surface roughness, and may be used for generalization and filtering.

INTRODUCTION

The problem of describing the forms of curves and surfaces has vexed researchers over the years. For example, a coastline is neither straight, nor circular, nor elliptic and therefore Euclidean lines cannot adequately describe most real world features. Imagine attempting to describe very rough or bumpy terrain in terms of classical geometry. An intriguing concept proposed by Mandelbrot (1967, 1977, 1982) is to use fractals to fill the void caused by the absence of suitable geometric representations. A fractal characterizes curves and surfaces in terms of their complexity by treating dimension as a continuum. Normally, dimension is an integer number (1 for curves, 2 for areas, and 3 for volumes); however, fractal dimensions may vary anywhere between 1 and 2 for a curve and 2 and 3 for a surface depending upon the irregularity of the form. Although individual fractals have been around since the 1900's, Mandelbrot was the first to recognize their applications outside of mathematics.

This paper discusses an algorithm, developed at the Ohio State University and implemented in an interactive setting, designed to measure the fracticality of a surface. It also presents results from examining three surfaces.

DEFINITION OF FRACTALS AND SELF-SIMILARITY

In Euclidean geometry every curve has a dimension of 1 and every plane has a dimension of 2. This is generally referred to as the topological dimension (D_t). These dimensions remain constant no matter how complex or irregular a curve or plane may be. For example, the Rocky Mountain area contains many irregularities, but the topological dimension remains 2.

In the fractal domain a curve's dimension may be between 1 and 2 according to its complexity. The more contorted a straight line becomes, the higher its fractal dimension. Similarly, a plane's dimension may be a non-integer value between 2 and 3. The fractal dimension for any curve or surface is denoted by (D) and within this framework: $D \geq D_t$. Mandelbrot (1977) proposes the following definition for a fractal: "A fractal will be defined as a set for which the Hausdorff-Besicovitch dimension strictly exceeds the topological dimension."

Central to the concept of fractals is the notion of self-similarity. Self-similarity means that for any curve or surface a portion of the curve or surface can be considered a reduced image of the whole. However, seldom in nature (crystals are one exception) does self-similarity occur and therefore a statistical form of self-similarity is often encountered. In other words, if a curve or surface is examined at any scale it will resemble the whole in a statistical sense; therefore, D will remain constant. Brownian motion is an excellent example of statistical self-similarity. Because of this principle, a curve can be decomposed into N=r nonoverlapping parts and each subsegment has a length of 1/r=1/N. Similarly, a unit square can be divided into $N=r^2$ squares, where the similarity ratio is $r(N) = 1/r = 1/N^{\frac{1}{2}}$. In either case the following equation applies:

$$D=\log N/\log (1/r) \tag{1}$$

and could be called the shape's similarity dimension. D can also be expressed as:

$$D = \log(N/N_o)/\log(\lambda_o/\lambda) \quad (2)$$

where λ_o and λ are two sampling intervals and N_o and N are the number of such intervals contained. If a curve resembles a straight line then when the sampling interval is halved, N doubles and the proportion equals 1. The majority of cartographic curves are not straight lines and therefore N will more than double causing D to be greater than 1. The principle of self-similarity is dismissed by Goodchild (1980), Hakanson (1978), and Scheidegger (1970). Hakanson, for example, points out the absurdity of postulating the validity of self-similarity down to the size of the pebbles on the coastline and at the molecular interstices of those pebbles. Goodchild demonstrates that although Richardson (1961) found the west coast of Britain to have a constant D of 1.25 over sampling intervals between 10 and 1000km., he found the east coast to vary between 1.15 and 1.31 for a similar sampling interval. This suggests that whatever created the irregularities on the coastline acted at specific scales.

DEVELOPMENT OF THE FRACTAL SURFACE ALGORITHM

The following algorithm is based upon the research performed by Goodchild (1980). He noted the earlier empirical work performed by Richardson (1961) and later extended by Mandelbrot (1967). Richardson measured the lengths of several frontiers by manually walking a pair of dividers along the outline so as to count the number of steps. The opening of the dividers (n) was fixed in advance and a fractional side was estimated at the end of the walk. The main purpose in this section of Richardson's research was to study the broad variation of Σn with n.

Richardson produced a scatterplot in which he plotted log total length against log step size for five land frontiers and a circle. Mandelbrot (1967) discovered a relationship between the slope (β) of the lines and fractal dimension (D). To Richardson the slope had no theoretical meaning, but to Mandelbrot it could be used as an estimate of 1-D, which leads to:

$$D = 1 - \beta \quad (3)$$

Goodchild computed the fractal dimensions of surfaces by constructing contour lines and calculating their lengths. He extended Hakanson's (1978) analysis in order to calculate D. In computing length, Hakanson counted the number of intersections between a grid and a coastline, and for a complex curve, it is possible that any even number of intersections could exist. Goodchild points out the scale of the map, which determines the size of the grid, is not strictly related to a sampling interval. In terms of D, it would be more appropriate to count the number of cells intersected instead of the number of intersections.

Goodchild estimated D for several self-similar surfaces generated by shear displacement (Mandelbrot 1975). He selected a contour line at a height equal to the mean of the minimum and maximum heights. Cells were aggregated into larger aggregates of 5x5, 7x7, 9x9, and up to 19x19, and classified them as above (black) or below (white) the contour line. To count the number of boundary cells or, in other words, compute the length of the contour line, as a function of aggregate size, a count was made of the number of black aggregates containing at least one white cell. This is where the cells are cut by the boundary. The log of the average number of boundary cells was then regressed against the log of the aggregate size. Goodchild calculated D for the contour line using Equation 3 and extending that same principle to a surface, the following equation is used

$$D=2-\beta \tag{4}$$

The following algorithm developed by Shelberg (1982) extends Goodchild's work in that D can be approximated for nonself-similar surfaces and is placed in an interactive setting. Because Goodchild's study only dealt with self-similar surfaces, he could select a contour line at a height equal to the mean of the minimum and maximum heights. A self-similar surface possesses the same statistical properties at all scales and at all locations; therefore, choosing the average contour line is a valid method. However, a surface may not be entirely self-similar and selecting the mean contour line may produce a false fractal dimension. For example, if a surface consisted of a plain and a mountain range, then selecting the average contour line would neglect both points on the plain and mountain range and bias the overall fracticality. Although two of the surfaces reported in this paper are self-similar, the algorithm is designed to analyze nonself-similar surfaces. Instead of choosing only the mean contour line, any number of contour lines, up to a maximum of 200 can be used to examine a surface.

All surfaces are in the form of a Digital Elevation Model (DEM) and the algorithm begins with the user inputting whether the fractal dimension will be determined by rows or columns. This option is provided so that a trend in the surface may be captured. The contour interval is then entered and the number of contour lines determined. Next, the maximum cell size, which is comparable to Richardson's opening of the pair of dividers, is input. The minimum allowable maximum cell size is 5 because any number less than 5 would leave the linear regression open to question.

For each contour line, cells are aggregated into larger aggregrates, up to the maximum cell size, and classified as above (black) or below (white) the contour line. To count the number of boundary cells, as a function of aggregate size, a count is made of the number of black aggregates containing at least one white cell. Figure 1 represents a 6x6 surface in which the contour line of 35 is denoted by the darker line on the surface.

The algorithm starts with the lowest contour line, in this case 35, and classifies each cell as white (1) or black (2); see Figure 2. Always beginning with a cell size of 1, it then compares each neighboring cell, along the columns, for boundary cells. For example, the comparison between (1,1) and (1,2); and (1,5) and (1,6) indicates a boundary cell is

35

6	14	24	39	40	48	60
5	17	27	30	42	47	59
4	16	26	28	36	52	65
3	14	25	32	43	50	62
2	15	27	30	34	48	61
1	12	22	31	33	44	58
	1	2	3	4	5	6

Figure 1. 6 x 6 surface with a contour line of 35.

6	1	1	2	2	2	2
5	1	1	1	2	2	2
4	1	1	1	2	2	2
3	1	1	1	2	2	2
2	1	1	1	1	2	2
1	1	1	1	1	2	2
	1	2	3	4	5	6

Figure 2. 6 x 6 surface where below contour line = 1 and above contour line = 2.

19

6	5	10	20	18	16	13
5	2	11	12	18	17	14
4	9	12	13	17	16	15
3	4	8	9	16	14	13
2	7	11	14	15	13	13
1	4	10	11	14	15	17
	1	2	3	4	5	6

Figure 3. 6 x 6 surface with a contour line of 19. A case where no boundary cells are encountered for cell sizes 3, 4 and 5.

not present. Conversely, the comparison between cells (1,4) and (1,5); and (6,2) and (6,3) indicates boundary cells exist and in these cases the number of boundary cells is incremented by 1. After the entire surface is examined using 1 cell size, the search begins again using a cell size of 2. Now cells (1,1) and (1,3); and (6,2) and (6,4) are compared and so on. This continues until the maximum cell size is reached. For each contour line the same process is repeated.

It is possible, for a given cell size, that no boundary cells are encountered. In Figure 3, this occurs when the cell size equals 3, 4 or 5. For example, the only possibility of encountering a boundary cell is in row 6 when the contour line equals 19. If the cell size equals 5, cells (1,1) and (1,6) are compared and no boundary cells are reported present. The results remain the same for the remaining comparisons within that row. For any given cell size, if no boundary cells are encountered, the contour line is eliminated from the calculations. Since the algorithm regresses log (average number of boundary cells) against log (cell size), without eliminating the contour line, but just deleting the cell size from the regression, could result in 1, 2 or 3 points used in defining the regression line and immediately poses obvious difficulties. Although, by eliminating some contour lines, the method neglects certain small portions of the surface, it accounts for greater variations than if only the mean contour line is selected. The strategy is to choose the lowest contour interval with the fewest contour lines being eliminated.

After counting the number of boundary cells per cell size for each contour line, a linear regression is performed. For every contour line, log (average number of boundary cells) is regressed against log (cell size). A D is computed for each contour line and extended to a surface by using Equation 4. A surface's fracticality is the average of all the D's for the included contour lines.

EXAMPLES AND RESULTS

Two of the three surfaces used as examples were generated by the shear displacement method and are statistically self-similar. Each is in the form of a 50x50 matrix and, on the CRT, are represented by a perspective view with hidden lines eliminated.

The first surface, in Figure 4, has a theoretical D of 2.2. Table 1 shows a rather large array of D values and the selection of numerous contour intervals is intended to display the stability of the algorithm. In an effort to preserve consistency among the results and not fall below the minimum acceptable maximum cell size, in the following examples, the maximum cell size is 5. In all cases, D is computed along the rows.

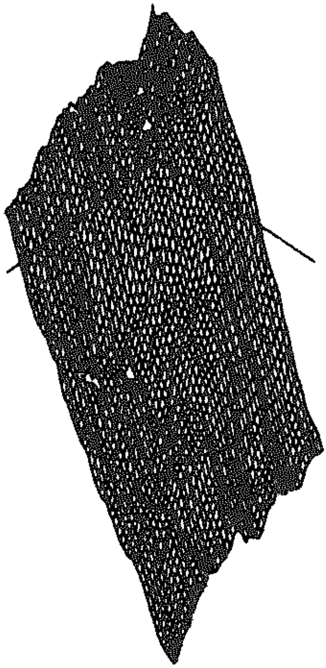


Figure 4. 50 x 50 surface with a theoretical D of 2.2. Minimum and maximum elevations are -503 and 269.

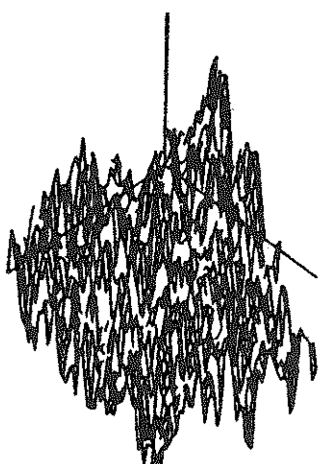


Figure 5. 50 x 50 surface with a theoretical D of 2.6. Minimum and maximum elevations are -42 and 41.

NOTE: Heights are unitless.

Contour Interval	Maximum Cell Size	D	No. of included lines	No. of lines not included
5	5	2.1915	146	8
10	5	2.1869	72	4
20	5	2.1732	36	2
22	5	2.2071	33	1
24	5	2.2000	30	1
30	5	2.1893	24	1
50	5	2.2365	14	0
100	5	2.2143	7	0
200	5	2.2268	3	0
385	5	2.1289	1	0

Table 1. Results from examining a 50x50 surface with a theoretical D of 2.2

A D of 2.1915 would be selected to represent the surface because of the large number of contour lines (146) used in the calculations versus only 8 contour lines being eliminated.

Figure 5 shows a 50x50 surface with a theoretical D of 2.6. Table 2 displays the results over a number of contour intervals.

Contour Interval	Maximum Cell Size	D	No. of included lines	No. of lines not included
2	5	2.6720	34	7
5	5	2.6671	14	2
7	5	2.6679	10	1
10	5	2.6290	7	0
20	5	2.6590	3	0
41	5	2.5985	1	0

Table 2. Results from examining a 50x50 surface with a theoretical D of 2.6

A D of 2.6720 would be selected because 34 contour lines are included in the calculations versus only 7 being eliminated.

The third surface is also 50x50, and is a Defense Mapping Agency's standard elevation matrix over a small portion of Southern Nevada. The minimum and maximum heights, in meters, are 2011 and 2376. Although the degree of self-similarity is not known, the D values, in Table 3, are extremely stable. A D of 2.1893 would be selected to best approximate the complexity of the surface. The actual surface is not shown because the hardcopy is illegible.

Contour Interval	Maximum Cell Size	D	No. of included lines	No. of lines not included
5	5	2.1893	66	6
15	5	2.1905	22	1
20	5	2.1707	16	1
150	5	2.1701	1	0

Table 3. Results from examining a "real-world" 50x50 surface.

SUMMARY AND CONCLUSIONS

The results demonstrate the stability of the algorithm in which, over a number of contour or sampling intervals, the fractal dimension remains reasonably constant. Other 50x50 self-similar surfaces which were examined, ranging in D from 2.1 to 2.9, produced similar results. By using a number of contour lines, the maximum variation in the complexity of the surface can be captured and therefore D can be closely approximated.

Potential applications of this algorithm include a new means to compute surface roughness, a means to measure the amount of surface generalization or filtering attained, and a method to store a compressed surface so that greater complexity can be introduced to match a target surface with a specific fractal dimension.

REFERENCES

Agterberg, F. P. (1982), "Recent Developments in Geomathematics," Geo-Processing, Vol 2, pp. 1-32.

Carpenter, L.C. (1981), "Computer Rendering of Fractal Curves and Surfaces," Unpublished Research Paper, Boeing Computer Services, Seattle, Washington.

Dutton, G. (1981), "Fractal Enhancement of Cartographic Line Detail," The American Cartographer, Vol 8, No. 1, pp. 23-40.

Goodchild, M. F. (1980), "Fractals and the Accuracy of Geographical Measures," Mathematical Geology, Vol. 12, No. 2, pp. 85-98.

Hakanson, L. (1978), "The Length of Closed Geomorphic Lines," Mathematical Geology, Vol. 10, No. 2, pp. 141-167.

Lam, N. S. (1980), Methods and Problems of Areal Interpolation, Ph.D. dissertation, The University of Western Ontario, London, Ontario.

Mandelbrot, B. B. (1967), "How Long is the Coast of Britain? Statistical Self-Similarity and Fractal Dimension." Science, Vol. 156, pp. 636-638.

Mandelbrot, B. B. (1975), "Stochastic Models of the Earth's Relief, the Shape and the Fractal Dimension of the Coastlines, and the Number-area Rule for Islands," Proceedings of the National Academy of Sciences, Vol. 72, pp. 3825-3828.

Mandelbrot, B. B. (1977), Fractals: Form, Chance and Dimension, Freeman, San Francisco, 365 pps.

Mandelbrot, B. B. (1982), The Fractal Geometry of Nature, Freeman, San Francisco, 460 pps.

Nordbeck, S. and B. Rystedt (1972), Computer Cartography, Studentlitteratur, Lund, Sweden, p. 39.

Richardson, L. F. (1961), "The Problem of Contiquity," General Systems Yearbook, Vol. 6, pp. 139-187.

Scheidegger, A. E. (1970), Theoretical Geomorphology (2nd ed.), Springer Verlag, New York, 333 pps.

Shelberg, M. C. (1982), The Development of a Curve and Surface Algorithm to Measure Fractal Dimensions and Presentation of Results, Unpublished Master's Research Paper, Dept. of Geography, The Ohio State University.

Shelberg, M. C., Moellering H., Lam, N. S. (1982), "Measuring the Fractal Dimensions of Empirical Cartographic Curves," Proc. AUTO-CARTO V/ISPRS IV Symposium, pp. 481-490.

Shelberg, M. C. and H. Moellering (1983), "IFAS: A Program to Measure Fractal Dimensions of Curves and Surfaces," Proceedings ACSM, 43, pp. 483-492.

AN ALGORITHM FOR VARIABLE-WIDTH FEATURE REPRESENTATION

Barbara Bonnette
TIGER System Staff
Geography Division
U.S. Bureau of the Census
Washington, D.C. 20233 U.S.A.

ABSTRACT

Automated cartography is becoming an increasingly important part of large geographic systems. The U.S. Bureau of the Census is preparing a totally integrated geographic computer-based system for use in the 1990 Census and beyond. One of the three major elements of the computer system is the requirement to prepare the maps used in the field operations. This means that efficient solutions to the problems of computerized cartography will have to be developed. One of the significant problems is that of drawing streets and other features as double lines from a data file which contains only centerline coordinates. Cartographically, double-line symbols can act as casings for feature names; the width of the casing can indicate importance, and the visual impression is more realistic.

This paper presents an efficient algorithm for plotting double-line symbols. It describes the reasons for its development, explains the supporting theory, describes the information necessary for the algorithm to operate, demonstrates applications as typified by census needs in field mapping, and presents weaknesses of the algorithm as a topic for future study. While it depends upon the geographic data base structure underlying the Census Bureau's system, the algorithm and the techniques involved in its use have general-purpose cartographic applications.

PROBLEM ANALYSIS: PRODUCTS TO PROCESSES

Digitizing the United Kingdom River Network 331
Roger Moore

Carte des sols et carte des terres agricoles 339
Jean-Phillippe Grelot

Cartographic and Attribute Data Base Creation for Planning Analysis Through GBF/DIME and Census Data Processing 348
Apollo Teng

Computer Mapping for Biomass Inventories 355
W.C. Moore and W. Chow

Flood Estimation in Europe: A Case Study of Applied Digital Cartography 365
Steven Wiltshire

A Geometric Mine Modelling System 374
Olli Karonen, Marku Tamminen, Penti Kerola, Matti Mitjonen and Esko Orivuori

Computer-Aided Design in Relation to Earthwork Computations .. 384
Stephen Mruk

Automated Data Base Capture for CADD 398
Richard Stover

Extended Graphical Plotting with PLANICOMP (Abstract) 408
Dierk Hobbie

MOSS - A State of the Union Address (Abstract) 409
Carl Reed

The Geographically Encoded String Structure (GESS) As Applied to the 1980 Census Digital County Boundary File (Abstract) ... 410
Roy Borgstede

Application of a Commercial Mapping/Data Management System to Forest Land Management in Maine (Abstract) 411
Daniel Boss

EPPL: An Innovative Grid Geographic Information System Language (Abstract) 412
Joseph Zastrow

A Three-Dimensional Model for Evaluating Potential Hydro-Geothermal Resource Utilization (Abstract) 413
Thomas Hodler

Some Problems in the Mapping of Geological Surfaces (Abstract) 414
Christopher Gold

DIGITIZING THE UNITED KINGDOM RIVER NETWORK

R.V. Moore B.Sc., M.Sc., D.I.C., M.I.W.E.S.
Institute of Hydrology
Crowmarsh Gifford, Wallingford, Oxon., OX10 8BB U.K.

ABSTRACT

Digitising the river network was originally proposed as a means of determining its length. Length was required for quantifying the extent of pollution and establishing a national river referencing system called the Hydrological Reference. Once proposed many other opportunities became apparent. Firstly, it made possible the automated production of simple river maps. Secondly, digitising gave access to the logical structure of the river network, that is the upstream and downstream relationship. In turn, this made possible the automated production of sophisticated river maps, it removed the technical obstacles to interrogating river related data banks by allowing users to point at an interactively generated map, and it opened up new ways of taking off river survey data. This paper reports on the progress and problems in digitising the river network of the U.K.

INTRODUCTION

In 1974 a bold project was started to design a system called the Water Archive[1] for the storage of scientific and technical water data. The aim was to provide an information system that matched and supported the objectives of the newly formed water authorities[2]. These had the then novel and idealistic purpose of managing the water cycle as a whole, and not, as is still often the case, as a number of separate and unrelated parts: water supply, sewage disposal, sewerage, pollution control, navigation or fisheries.

At an early stage in the archive's design, a number of key ways of cross-relating data were identified, and, for the purposes of this paper, the most important were geographical (x,y and z co-ordinates) and hydrological position (position in relation to the river network)*. Rivers, by their nature, form the focal point of a water authority's responsibilities. Their uses are many and conflicting, and it is the task of an authority to resolve and police the competing requirements. To do this, knowledge of the upstream and downstream relationship between the people, properties, structures, abstractions, discharges and observations involved is vital. It was therefore agreed that the relationship should be built into the archive system. This was achieved by adapting the American River Mile Index to create a new system called the Hydrological Reference (HR). In essence, it is a hierarchical numbering system based on river distance.

At the same time as the Water Archive was getting under way, a national survey of river pollution[3] was being conducted. It reported on the state of the rivers, classifying them by degree of pollution and quantifying

---

* As an aside, it is interesting to note that they were made the main referencing systems as much for the fact that they were apolitical and unaffected by administrative change, as for their more obvious benefits.

problems in terms of the affected stream length. A main part of the report was a highly acclaimed set of maps, showing the rivers coloured according to their state of pollution and with a width proportional to their flow. The river lengths had been assessed many years before and lacked consistency. Map production was by manual means, which was tedious, time consuming and error prone. It was also so huge an undertaking as to threaten the future viability of the survey. New simpler, quicker and less error-prone techniques were needed.

As the Water Archive System developed and began making information available to a wider range of people, many of whom had not previously encountered computers, so the pressure for better and more imaginative output increased, particularly for graphical and map output. In these instances the rivers were only needed as background information, but it was important to be able to control the detail with which they were shown both in terms of extent and generalisation.

The usefulness of being able to select and order data by their position on the river system was quickly confirmed once the system became operational, though the means by which the facility was made available was not popular. What was required was for users to be able to quote a national grid reference (NGR), something with which everyone in the U.K. is familiar, and for the system to be able to use that to locate the appropriate river, and then deduce what is upstream or downstream. The solution to this problem opened up many new possibilities of which two of particular interest are (i) allowing users to create and then point at a map as a means of interrogating a databank, and (ii) pointing at a map as a means of entering data about rivers. In both cases the map would either be placed on a digitising table or generated on a graphics screen.

Rivers once digitised form an immensely valuable base of data for scientific analysis, and relationships will be able to be explored that have previously proved impracticable, for example, the relationship between river geometry and flooding.

The objectives of digitising the river network therefore emerge as 1) providing a reliable and consistent estimate of length; 2) recording the structure of the network; 3) making it possible to locate the position of a point in the network from a grid reference; 4) being able to determine what is upstream or downstream of it; 5) being able to compile specialist river maps and more conventional maps in which the rivers are shown only as background at an appropriate level of detail; 6) to open up new ways of interrogating databases and collecting survey data; and finally 7) to provide a base of data for scientific analysis.

This paper describes the work to date, summarises the conclusions and reports on the way ahead.

PROGRESS AND PROBLEMS

An outline of the Water Archive System

Before relating the project history, it is useful to describe briefly the Water Archive System which has provided a vehicle for much of the work.

The design of the system was influenced by three things, the 1973 U.K. Water Act, the STORET water quality system[4] of the American Environmental Protection Agency and the Canadian NAQUADAT water quality system[5]. Its capacity to handle a wide range of data stems from the requirement of the Act to manage the water cycle as a whole. Its flexibility derives from the dictionary ideas of the two North American Systems.

The Water Archive is a highly generalised system for the storage of all types of technical and scientific data of interest to the water industry. It is designed round two types of data:-

data about places - feature data;
time series data.

Examples of features are dams, bridges, reservoirs, raingauges, sampling points, sewage works, pumping stations, and offices. The limits of river characteristics are also treated as features, examples being mooring rights, fishing rights, the limits of pollution and the extent of fish distribution[6]. Feature data describe the time invariant attributes of features. Examples of time series are quantities of water abstracted and sewage discharged, chemical concentrations, rainfall, evaporation, river flows, power consumption, bacterial counts and fish kills. The system is entirely general and the user defines the types of feature and time series in a set of dictionaries, which have become a means for standardising coding and nomenclature.

The system is designed as a series of linked sub-systems. At the hub is the feature sub-system and it is here that all place data are recorded and spatially cross-referenced, e.g. by grid or hydrological reference. It is also here that the structure of the river network is held as a set of features describing the mouths, confluences and sources of rivers.

Co-ordinate data are held in a boundary file. Originally this file was intended to store boundaries for point-in-polygon searches. However, its simple and efficient indexed design has meant that it can be used for all types of co-ordinate data.

Hydrological References

Returning to the problem of the rivers, a Hydrological Reference defines the position of a point on the river network, and given two references it is possible to determine whether one is upstream or downstream of the other, or if they are on separate parts of the river system. The logic uses the relative distances of the two points from the sea, combined with a knowledge of the structure of the network. A Hydrological Reference does not contain all the information within itself, and the structure of the network is held within the databank by storing all the nodal points, i.e. the confluences and sources.

Trial digitising in the South West Water Authority

Although at the time Hydrological References were set up the only feasible way to measure the river length was by hand, it was recognised that the best way would be to digitise the rivers. Therefore a trial project was put in hand with the Ordnance Survey (O.S.), the U.K. national mapping agency, to study the required scale, accuracy, data volumes, data structuring and costs. The area chosen was a 40 x 40 km square in the south west of England. Defining the extent of the network to be digitised was a problem. It was and has continued to be difficult to devise any scientific criteria for deciding that a river is relevant or irrelevant. Lacking any other criteria, the network of interest was defined to be those streams shown on the 1:50,000 series maps of the Ordnance Survey together with any additional streams of interest to the water authorities. The results of the trial were encouraging. A sustained accuracy of ± ½ a line width (0.075mm) was achieved. The amount of data was 68,000 coordinate pairs and 3000 line segments, a segment being the river between two confluences. About 50% of the co-ordinates could be dropped without significant loss of accuracy.

The scale of 1:50,000 was found to be satisfactory in relation to the purposes of the data. Costs for the project were £1,400 (1976 prices) and production costs were estimated at £700 per 40 x 40 km sheet of which there are 204 for the U.K.

Unfortunately, it was not possible to continue as a result of government cut backs, and so attention was turned to the production of an experimental system whose objectives were the production by computer of the River Quality Survey (RQS) Maps and the production of maps directly from data in the Water Archive System.

Initial software development

The RQS maps show the river with a colour proportional to the degree of pollution and a width proportional to the flow. There were two options for organising the production of the maps. The first was to set flags along the lists of digitised co-ordinates to indicate the level of pollution or flow. This was a very attractive proposal because of its simplicity. However, investigation soon showed it to have very severe long term drawbacks. Principal of these was that a flagging system, once adopted, would be applied to all sorts of other data e.g. fish distribution, and a facility originally intended for co-ordinate information would end up holding all manner of data for which it was never intended, and the demands of which it would not be able to meet. The approach adopted, therefore, was to say that co-ordinate data are a perfectly valid set of data in their own right, and should be stored separately. By the same argument, the quality data are also a complete and independent set of data, and have in fact been stored as a set of features recording the points of change in the pollution class. This intuitive solution is of course confirmed by any formal method of data analysis. The link between the two data sets has been established by labelling both with HR's. Out of this approach, a general solution has been developed where the style in which rivers are depicted can be controlled by an independent and separate data set.

The principal result of this exploratory work has been a robust batch mapping system, the maps being considered here requiring too much data and processing for real time interactive production. In outline, the system comprises an interactive prompt by which the user specifies the required map. The prompt writes the job control language to invoke the appropriate sub-systems to retrieve and process the necessary data. Each sub-system appends the data to be plotted to a file. At the end the file is sorted by pen width and colour, and merged with any required pre-existing files which might contain, for instance, a special version of the coast or an overlay of selected raingauges. The sorted file is then passed to the plotting program, which has as its control parameters, inter-alia, the scale and area to be shown.

The North West Water Authority Pilot Project

The successful demonstration of the experimental mapping system, renewed interest in digitising the rivers, and a pilot project to digitise the rivers of the North West Water Authority (NWWA) was undertaken by NWWA and Department of the Environment (DOE). Its objectives were to digitise the rivers of NWWA, to store the co-ordinate data in a structured manner in the Water Archive System, to replace all the old manually derived river distances by digitally derived distances and finally to make possible the derivation of an HR from a national grid reference (NGR). By products seen as emerging from the work were a digital data editor, an interactive map compiler, an interrogable map and new data collection techniques for the River Quality Survey. Data and project management were also of interest.

The objectives were achieved by a three stage process. Stage I comprised: checking TRB features a pre-existing list of the mouths, confluences and sources; numbering TRB features; producing a schematic plot of rivers; map marking; digitising; validating and finally linking digital data to TRB features. The second stage involved replacing manually derived distances in Hydrological References, and the final stage was to compile an NGR to HR conversion table.

Water Authorities who use the Water Archive System have lists of all mouths, confluences and sources of their rivers of interest. These are collectively referred to in the Water Archive as TRB (tributary) features, and each TRB has pointers to the next TRBs immediately upstream. The first step was to check rigorously the correctness of the TRB network. When that was achieved, the TRBs were numbered. The numbers assigned were for use as a pointer system to link the TRBs to the digital coordinate data.

A schematic plot was then produced to check that the TRB features were correctly located, and to identify all the streams to be digitised. The schematic plots were used as a guide to the third and crucial stage, map marking. Correcting errors in digital data is a messy, time consuming process, and therefore good map preparation is essential. Map marking has three functions: identifying to the digitiser the rivers to be digitised, filling in gaps in the river network, e.g. rivers in culvert or under bridges, and adding identifying stretch codes.

The maps used were new 1:50,000 paper copies and the rivers were identified by over marking them with a see through felt tip pen. After digitising, the data were validated in two steps. The first step concerned map skew and stretch correction. In the final step, the ends of all lines meeting at confluences were matched exactly and the grid references of the TRB features were adjusted to coincide with the digitised confluence. Digitised and manually derived stream lengths were compared. This was an iterative procedure and each pass produced a check plot with all errors identified in colour. A blue map indicated a completely valid data set.

The second stage, for which the software has been written but not yet applied, concerns the replacement of all manually derived stream length estimates used for Hydrological References, by digitally derived lengths. In principle, this is a relatively straightforward process. In practice, it is likely to be more difficult, firstly, because of the sheer magnitude of the operation and secondly, the need to provide very good security as a guard against machine failure during the run. Hydrological References are the central referencing system of the Water Archive System and their logical correctness is vital.

The third stage involves the compilation of a table to enable grid references to be converted to Hydrological References. A 1 x 1 km grid is set up for the country. Each square is indexed by the grid reference of its south west corner. For each square a list is compiled of all the stretch reference numbers for all the lines that cross the grid square. In fact, the list is for a slightly larger square to allow for rivers that run along grid square edges. Implicit in any grid reference is the reference of the SW corner of the grid square that contains it. It is therefore easy to get quickly to the list of stretches that might be near the given grid reference. It is then a simple matter to scan the lines in the square and find by interpolation the point on the river nearest the given grid reference. This table is the key to making it easy for user to access the rivers and related data, and is vital if interactive systems are to be considered.

The digitising work for the North West Authority area is now complete and only stage II remains; all the remaining objectives of the project have been achieved. An interactive editor for correcting digital data has been prepared and used. The ideas in the editor have been developed into an interactive map compiler which allows the user to sit at a screen and build a map. The off-line mapping system has been further developed to handle the volumes of data that arise in generating River Quality Survey maps for major rivers. The last piece of work concerned developing software for producing the maps proposed for the European Economic Community Fish Directive. These maps are similar in principle to the River Quality Survey Maps. The important step forward has lain in the method of compilation. It is now only necessary to call up the relevant river on the screen and point with a cursor to the place at which the designation of the river changes from coarse fish to salmon. Once this has been done a coloured map of fish distribution can be prepared and river lengths compiled.

The Yorkshire Water Authority

During the NWWA/DOE project much was learnt particularly about efficient procedures, accuracy and errors. New techniques of digitising were emerging, particularly the Laserscan Fastrack system. A second pilot project was therefore set up with the Yorkshire Water Authority (YWA). The project objectives are to assess the Fastrack system, to assess the automatic structuring of unstructured data, to simplify map marking, to assess the data volumes that would result from following the latest O.S. accuracy standards and to assess the costs.

The O.S. was contracted to undertake the work. There were a number of reasons for seeking to work with the O.S. as opposed to any other contractor. Firstly, there was the long term maintenance of the data, a role for which the O.S. exists. Next there was the fact that once the data were available, they would almost certainly be of interest outside the water industry, they would be a national asset, and once again the O.S. would be the best organisation to make it availahle. It was important to the Water Industry that the digitised representation of the rivers should remain compatible with the published Ordnance Survey maps. Co-operating with the O.S. and accepting their accuracy standards waa probably the most practical way of ensuring that the digital database that the O.S. is building meets not only their map production needs hut also the Water Industry's need for data.

The specification agreed with the Ordnance Survey may be summarised as a digitising scale of 1:50,000, all the lines on the 'blue plate' of water features, lines coded to indicate, interalia, river centre line, river bank, lake shore, and coast, all gaps in line work filled on rivers of interest to the water authority, accuracy defined as no point on a digitised line to depart from the original map hy more than 0.075 mm, and exact node matching.

The project procedure comprised three steps: map marking by YWA, digitising by O.S., and structuring by DOE/IH. Map marking was greatly simplified. The need to numher stretches at digitising time was eliminated, thus removing the major opportunity for error.

It was hoped to achieve a major reduction in production time and ease the data structuring problems by using laser line following equipment, and a new piece of software for automatic junction recognition. Unfortunately the Ordnance Survey's preliminary research did not correctly predict the proportion of rivers shown double-sided and this, compounded by software problems, led to a serious rise in costs to the point at which the project had to be stopped. However, every effort is being made to rectify the situation and bring the cost back to an economic level. On the positive side, a breakdown of the quantities of river data by line type now exists that is far more detailed than any previous estimate.

Data storage proposals

Although the Water Archive System has been used to store the digital river data acquired so far, it was never intended for holding large volumes of co-ordinate data. Accordingly, a study has been made as to how the Water Archive System design might be revised to accommodate such data. In reviewing current and proposed systems, it was concluded that the main point yet to be achieved was a simplicity of concept. It was also realised that the terms, 'automated cartography' and 'digital mapping', while useful, are also misleading. They divert attention from the central problem, which is the orderly storage of information, to what ought to be a relatively peripheral one, the mechanics of presenting the data.[7]

Study showed that the requirement was simple and conventional. In essence, the user saw the world as composed of features qualified by attributes. In most cases, digitised lines were merely special cases of attributes. A variety of ways of implementing the system were considered, including a number of databases. At present, it is suggested that five indexed sequential files be used. The principle file is the feature file which records for each feature any number of dated attribute lists. Each feature can be cross-referenced to any other feature and to any set of lines. This allows relationships to be established between features. For example, there can be a feature that is the entire River Thames, and there can be separate features for each individual stretch of the Thames. The other four files contain the digital data, an inverted file of attribute values, the node link structure of the data and a register of codes and titles.

CONCLUSIONS AND FUTURE WORK

All the major technical obstacles to obtaining digital information on rivers have been overcome. Sound new methods exist for producing good clear maps, cheaply and in a reasonable span of time at an appropriate quality. The only limit to the development and exploitation of these systems is the user's imagination. New techniques of collecting survey data have been devised and now desirable surveys whose future was in jeopardy on financial grounds could continue. Regrettably, modern digitising techniques have not proved successful, and a return to manual digitising run on a cottage industry basis is being considered as a means of keeping costs within bounds. The task currently in hand is the design of a new storage system.

ACKNOWLEDGEMENTS

The author gratefully acknowledges the permission of the Director of the Institute of Hydrology to publish this paper. He would also like to thank all those involved in the project and who have supported it, in particular, Mr J Tyson and Mrs P Woods of the North West Water Authority, and Mr D Rodda and Dr R Otter of the Department of the Environment.

REFERENCES

1. Rodda, D.W.C., Moore, R.V. and Liddament, J.L. 1978, The Water Archive - Phase 2, A Functional Description, Department of the Environment, Water Data Unit, Reading.

2. The Water Act 1973, 1973, H.M.S.O., London.

3. Department of the Environment, 1978, The River Pollution Survey of England and Wales 1975, H.M.S.O., London.

4. Environmental Protection Agency, The Storet Users' Manual, E.P.A., Washington, D.C.

5. Daymo, A. et al, The Naquadat User's Manual, Environment Canada, Ottawa.

6. Commission of the European Community, 1979, Directive 78 659 on the Quality of fresh water needing protection or improvement in order to support fish life, Commission of the European Community, Brussels.

7. Bickmore, D. 1982, ed. Perspectives in the Alternative Cartography, Cartographica, Vol 19, Number 2, Monograph 28.

CARTE DES SOLS ET CARTE DES TERRES AGRICOLES

Jean-Philippe Grelot
Institut Géographique National
2, avenue Pasteur
94160 SAINT MANDE-FRANCE

RESUME

Parallèlement à l'accroissement des phases automatiques de la rédaction de la Carte des Sols de la Région Centre, une réflexion a été menée par les cartographes de l'Institut Géographique National-France et les pédologues responsables de ce programme en vue d'une meilleure utilisation des données numérisées.

C'est ainsi qu'il a été décidé d'entreprendre la confrontation des données pédologiques, définissant une aptitude agricole des sols, avec l'occupation du sol observée sur photographies aériennes.

La possibilité de dresser une carte de la valeur des terres agricoles, carte définie en application de la loi d'orientation agricole de juillet 1980, a modifié cet axe de recherche. Il apparaissait en effet que les données pédologiques numérisées pour établir la Carte des Sols étaient l'élément de base pour dresser cette carte des terres agricoles, et que des données auxiliaires (occupation des sols, pentes...) pourraient les compléter pour créer, au delà de la carte proprement dite, une base de données.

Il ne restait plus alors qu'à définir les données à recueillir et modéliser leur traitement. Les développements informatiques ont consisté à écrire le programme correspondant, mais surtout à constituer un logiciel permettant de rassembler des données thématiques zonales en une base de données aux applications multiples.

Le champ d'expérimentation a été la feuille de Léré. Le résultat escompté a été obtenu, tandis que se sont révélées des potentialités d'exploitation des données extrêmement prometteuses.

INTRODUCTION

Engagé en 1980, le programme de rédaction cartographique de la Carte des Sols de la Région Centre n'a cessé de se développer. Deux idées directrices peuvent résumer cette évolution : automatisation accrue de la cartographie d'une part, utilisation de la base de données numérisées d'autre part.

Cette expérience peut à maints égards être considérée comme une opération pilote, dans laquelle les développements des exploitations et des logiciels de cartographie thématique s'enchainent dans une synergie particulièrement efficace. Des horizons nouveaux s'ouvrent et des méthodologies nouvelles s'acquièrent, au plus grand bénéfice des deux communautés-pédologues et cartographes - et à travers elles, de tous les utilisateurs.

Un des fruits de ces travaux est l'obtention d'une carte de la valeur des terres agricoles. La Carte des Sols et ses thématiques dérivées sont complétées par la numérisation de données auxiliaires. Une base de données commune est constituée, ce qui permet de multiples exploitations. Au-delà des résultats acquis, des développements sont d'ores et déjà envisagés.

LE PROCESSUS DE NUMERISATION ET DE REDACTION

La première phase du processus cartographique consiste à numériser le fond plani-oro-hydrographique, orienté selon le système Lambert, en lui faisant subir éventuellement quelques corrections géométriques pour compenser les déformations. Cette opération offre deux avantages : on dispose d'un fond reproductible sans altération que l'on peut assembler parfaitement avec les feuilles voisines, et on pourra combiner directement cette image numérique avec celle de la planche thématique du noir, sans opération de photogravure.

Les différentes maquettes sont rédigées en s'appuyant sur ce fond. Les contours des zones sont dessinés sur une planche séparée des informations thématiques rendues par des codes ou des couleurs. Cette séparation correspond aux deux étapes de la saisie, dont la première traite les seules informations géographiques (localisation des zones homogènes) et la seconde les renseignements thématiques (contenu de ces zones).

La numérisation des contours se fait sur un scanneur à balayage. Des corrections géométriques sont effectuées, de la même façon que pour le fond topographique : on pourra donc parfaitement superposer plusieurs thèmes ou assembler des feuilles voisines. Les contours sont ensuite automatiquement réduits à leur squelette, puis diverses corrections sont opérées interactivement sur un écran graphique couleur. Enfin, un programme identifie tous les points d'une même zone délimitée par un contour fermé et leur attribue un code numérique unique.

La seconde étape consiste à affecter les attributs thématiques à chacune des zones identifiées par ce code. Cette codification est effectuée sur une table à numériser, tandis qu'un programme interactif guide l'opérateur en lui signalant sur un terminal alphanumérique l'enchainement des opérations, la localisation des zones à coder et les erreurs décelées (par exemple, pointés contradictoires). Une double codification permet de réduire considérablement le taux d'erreur.

Les poncifs et les trames ont été dessinés sur un écran graphique. Un fichier de légende a été numérisé, qui comporte les caissons et toutes les indications (titres et sous-titres) communes à l'ensemble des cartes de la série. Cette opération permet de réaliser des économies susbtantielles de coûts et de délais.

La restitution quadrichroromique est faite au moyen d'une caméra à laser. Quelques compléments manuels peuvent être apportés avant de confectionner une épreuve d'essai par procédé Cromalin.

Les fichiers sont archivés après corrections éventuelles. Les données numériques sont donc constamment disponibles pour tout traitement, qu'il soit de nature cartographique ou non.

LA CARTE DES SOLS

Le profil de chaque unité de sol figurant sur la carte au 1:50 000 est reconstitué par cinq éléments graphiques qui traduisent les informations fondamentales d'ordre pédologique et agronomique nécessaires à tout utilisateur et qui sont :

- le développement du profil d'après la classification française, représenté par une couleur ;
- la texture du profil, schématisée par une surcharge noire de points, de tiretés ou de traits ;
- l'intensité de la stagnation de l'eau, en surcharge de tirets bleus ;
- la charge caillouteuse, représentée par des symboles géométriques noirs empruntés à la schématisation habituelle des géologues ;
- la nature et la profondeur d'apparition du substrat ou d'une discontinuité lithologique, ainsi que son altération et/ou sa texture éventuelles, traduites par des caractères typographiques.

Cinq feuilles sont dès à présent éditées ou en cours d'impression. La principale amélioration apportée depuis le traitement de la première feuille (Léré) est le positionnement automatique des caractères typographiques symbolisant le substrat : cette méthode a été utilisée pour la feuille de St Amand Mont-Rond.

LES CARTES DERIVEES DE LA CARTE DES SOLS

Quatre cartes dérivées (textures superficielles, contraintes liées à l'excès d'eau, réserves utiles potentielles, aptitudes agricoles des sols et facteurs limitants) sont éditées à l'échelle de 1:100 000 de manière entièrement automatique après la validation des fichiers de la Carte des Sols.

Le fond topographique et la légende ont été numérisés, toujours pour réduire les interventions manuelles et les opérations de photogravure, ce qui accélère la fabrication de ces cartes.

Pour chacune d'elles, les pédologues ont établi un modèle utilisant les différentes informations numérisées. Le programme écrit à partir de ce modèle permet naturellement la mise en teintes des cartes dérivées, mais également la sélection automatique des contours de zones à partir de la totalité des contours de la carte de base. De même, il est possible de positionner automatiquement un symbole au centre de chacune de ces nouvelles zones.

La carte des textures superficielles reprend le thème texture de la Carte des Sols, en remplaçant les poncifs lignés ou tiretés noirs par des couleurs.

La carte des contraintes liées à l'excès d'eau est obtenue par croisement des thèmes type de sol (sélection des sols hydromorphes minéraux) et intensité de la stagnation de l'eau.

La réserve utile potentielle en eau sur un mètre de profondeur est calculée pour chaque zone à partir des indications de texture, de charge caillouteuse et de substrat. On procède ensuite à un regroupement de ces valeurs en neuf classes.

La totalité des informations numérisées et la réserve utile potentielle en eau sont utilisées pour calculer une aptitude agricole des terres, exprimée par une valeur de 0 à 100 points. La carte est établie par un regroupement de ces valeurs en dix classes. Elle est complétée par l'indication au moyen d'un symbole typographique du facteur limitant prépondérant pour chaque zone ou regroupement de zones de même classe d'aptitude obéissant à la même contrainte majeure.

Ces données calculées (réserve utile, aptitude, facteurs limitants) sont archivées, en complément des données pédologiques numérisées.

LES DONNEES AUXILIAIRES

Dès l'édition des premières feuilles et à la suite des réflexions ayant conduit à la production des cartes thématiques dérivées, s'est manifestée l'idée d'étudier l'adéquation de l'usage du sol et de ses caractéristiques agricoles telles que l'on peut les appréhender à partir des données pédologiques. Une action de recherche a été entreprise conjointement par l'I.G.N., l'Institut National de la Recherche Agronomique (I.N.R.A.) et les Chambres d'Agriculture du Cher et de l'Indre, pour définir et recueillir les informations nécesssires, puis élaborer une modélisation et développer un logiciel permettant le traitement de l'ensemble de ces données.

En outre, la promulgation de la loi d'orientation agricole de juillet 1980 qui prévoit, entre autres dispositions, l'établissement d'une carte de la valeur des terres agricoles, a accru l'intérêt suscité par cette démarche : la Carte des Sols pouvait en effet être le support indispensable à la réalisation de cette carte.

La feuille de Léré a été choisie comme champ d'expérimentation. Plusieurs types de données ont été collectés et numérisés. On peut les classer en données physiques (occupation du sol, pente) et données d'aménagement (remembrement, drainage, irrigation).

L'occupation du sol a été dressée par photo-interprétation d'une prise de vues aériennes au 1:20 000 effectuée en juin 1982. Quatorze modes d'occupation ont été retenus : terres labourées, céréales, maïs, colza, protéagineux, vigne, horticulture, prairie artificielle, prairie pérenne, bois feuillu, bois résineux, broussailles et landes, eau, habitat et urbanisation. Les haies ont également été relevées et numérisées. A partir de cette carte d'occupation du sol et en tenant compte de critères annexes comme la zone d'appellation contrôlée "Sancerre", on a pu délimiter les agrosystèmes.

Le modèle numérique de terrain de l'I.G.N. a permis de dériver automatiquement un modèle numérique de pentes, dont on a dressé une cartographie.

Les aménagement agricoles ont été recensés. Les zones de drainage et les limites de communes, avec indication des communes remembrées, ont été numérisées.

Toutes ces informations ont été mises sous forme de fichiers maillés. Des sorties de contrôle ont été effectuées au fur et à mesure de la numérisation et de la codification.

LE TRAITEMENT DES DONNEES

A ce stade intervient de manière déterminante la forme sous laquelle est décrit l'espace. Le fichier maillé le découpe en une grille régulière dont chaque point est un carré de 0,1 millimètre de côté, ce qui représente 25 mètres carrés à l'échelle de 1:50 000. Ces points sont rangés séquentiellement, ligne après ligne, par indication du code numérique attribué à la zone homogène contenant le point considéré ; les coordonnées de localisation des points ne figurent pas à proprement parler, elles sont reconstituées à partir du rang qu'occupe le point dans le fichier. On peut donc très facilement parcourir plusieurs fichiers simultanément, et recueillir ainsi les différentes informations qu'ils décrivent.

Ceci est l'idée générale du processus ; les choses sont légèrement plus complexes dans la pratique, en raison des opérations qu'il a fallu faire pour compacter les fichiers afin de réduire le volume de stockage. Mais cette méthode demeure infiniment plus simple que celle requise pour effectuer un traitement similaire sur des fichiers décrits en mode vectoriel, où chaque zone est décrite par les coordonnées de la ligne la délimitant.

Dans l'exemple qui nous intéresse ici, une base de données a été constituée avec les fichiers de la Carte des Sols (et par voie de conséquence, de ses thématiques dérivées ainsi que des facteurs limitants), de l'occupation du sol, des agro-systèmes, des zones de drainage, des pentes et des limites communales. A partir de ces données, un modèle permet de calculer en chaque point une valeur agricole ; ce modèle est essentiellement une adaptation de celui utilisé pour dresser la carte thématique des aptitudes agricoles à partir de la Carte des Sols, dont on voit clairement le rôle primordial qu'elle occupe dans cette opération.

UNE CARTE DE LA VALEUR DES TERRES AGRICOLES

La valeur agricole calculée vient naturellement enrichir la base de données. On peut alors procéder à diverses exploitations.

Tout d'abord, on peut noter que l'information (recueillie ou résultant d'un traitement) est conservée sans dégradation par transcription sur une bande magnétique et duplication de celle-ci. Ceci peut paraître accessoire à première vue, mais devient fondamental si l'on désire réutiliser les données pour une exploitation future ou une mise à jour.

Un résultat que l'on peut obtenir, et qui n'est pas cartographique, est un fichier ou un tableau de données statistiques, correspondant à un planimétrage automatique. Il peut s'agir de surfaces totales pour la feuille entière ou bien par entité administrative (commune, région agricole), ou encore par région naturelle ou agro-système ; mais on peut décomposer ces données synthétiques en fournissant, par exemple, le mode d'occupation du sol (en hectares- tableau 1 - ou en pourcentage de la surface agricole utilisée - tableau 2 -) par classe d'aptitude (et éventuellement à l'intérieur de celle-ci, par facteur limitant) et par agro-système : c'est dire la richesse que recèle la base de données.

Tableau 1 - Répartition du mode d'occupation par classe de productivité (Surfaces en hectares)

Agro-système	Mode d'occupation	classe de productivité						Total
		1	2	3	4	5	6	
Montagne des Marnes et Butte de Jars	Prairie pérenne	-	144	176	117	18		455
	Prairie artificielle		22	43	6	6		77
	Céréales	2	243	238	138	26		647
	Maïs		36	11	18	1		66
	Terre labourée				3			3
	Surface agric. util.	2	445	468	282	51		1248
	Surface totale	2	485	505	305	59		1356
Sancerrois	Prairie pérenne	5	38	16	6	10	10	85
	Prairie artificielle	2		4	1		12	19
	Céréales	4	59	25	1	4	27	120
	Colza	-	3	4	-		7	14
	Maïs			2			2	4
	Terre labourée		3	8	1	3	1	16
	Vigne	13	160	512	3	73	92	853
	Surface agric. util.	24	263	571	12	90	151	1111
	Surface totale	31	289	604	18	144	175	1261

Tableau 2 - Répartition du mode d'occupation en pourcentage de la surface agricole utilisée, par classe de productivité -

en %

Agro-système	Mode d'occupation	Classe de productivité						Total
		1	2	3	4	5	6	
Montagne des Marnes	Prairie pérenne	24	32	38	42	36		35,5
	Prairie artificielle		5	9	2	11		6,2
et Butte de Jars	Céréales	76	55	51	49	51		51,8
	Maïs		8	2	6	2		5,3
	Terre labourée				1			0,2
Sancerrois	Prairie pérenne	20	14	3	48	11	6	7,6
	Prairie artificielle	7		1	11		8	1,7
	Céréales	15	22	4	8	5	18	10,8
	Colza	1	1	1	1		5	1,2
	Maïs			-			2	0,4
	Terre labourée		1	1	6	3	-	1,4
	Vigne	57	61	90	26	81	61	76,7

La mise à jour se fait, quelles que soient les évolutions, par saisie séparée des modifications et incorporation automatique dans le fichier thématique correspondant, puis dans la base de données elle-même.

Les sorties cartographiques peuvent revêtir plusieurs formes, selon leur finalité et la complexité des informations qu'elles expriment. Ainsi, des sorties monochromes (donc peu onéreuses) seront réalisées pour des phases d'étude ou de contrôle. On peut produire d'autres cartes dérivées par sélection et/ou combinaison de plusieurs thèmes. Ensuite, on pourra établir une carte de la valeur des terres agricoles sur l'ensemble de la feuille, ou bien en occultant les zones non agricoles (forêts, urbanisation, eau).

Cette carte prend en compte les données physiques du milieu, qui sont des éléments relativement stables et quantifiables. Les données économiques, plus conjoncturelles, ont été délibérement exclues du modèle de traitement. Leur examen peut néanmoins amener à modifier le classement résultant de la valeur agricole calculée. La carte établie à partir de ces seules valeurs calculées apparaît ainsi comme une maquette synthétisant les données physiques, et les modifications explicitées par les facteurs économiques interviennent comme des corrections avant l'édition de la carte définitive. Une alternative peut être proposée, en conservant la carte "calculée" et en exposant les facteurs économiques et leurs conséquences dans la notice accompagnant la carte.

LES DEVELOPPEMENTS

Les essais menés jusqu'à présent ont donné un aperçu de ce que permettait la méthode de combinaison des fichiers maillés en une base de données. Ils ont montré également que le champ d'expérimentation était très ouvert, en ce sens que de nombreux modèles peuvent être testés, sans que la technique impose des contraintes rédhibitoires.

Il s'agit à présent de reprendre l'idée initiale qui a présidé à cette démarche : comparer l'occupation du sol et ses capacités agricoles découlant des données pédologiques. Néanmoins, certaines données auxiliaires numérisées, comme les pentes ou le drainage, peuvent maintenant être prises en compte. Deux actions sont prévues : la première étudiera la répartition de l'occupation du sol par type de sol et inversement, les types de sols sur lesquels se développe un mode d'occupation donné ; la seconde cherchera à connaître l'évolution de l'usage du sol sur une longue période (plus de 20 ans), et son incidence sur les espaces fourragers, la jachère, les agro-systèmes et le parcellaire.

A cet aspect thématique de l'exploitation de la base de données, il faut ajouter les développements de logiciel à entreprendre. En effet, s'il a été prouvé que la méthode était praticable, certains problèmes se sont posés, comme la mise en accord des différents fichiers (ajustements géométriques, limites communes) ou l'allègement de la procédure de mise à jour. Des améliorations sensibles peuvent être apportées, qui diminueront notablement le temps de calcul nécessaire pour les applications décrites dans cet exposé.

CONCLUSION

Outre la production d'une carte de la valeur des terres agricoles, la technique utilisée montre l'intérêt de la cartographie thématique numérique appliquée aux données d'inventaires, telles celles de la Carte des Sols. Loin d'être figées, ces données peuvent être sollicitées à chaque instant et être confrontées à d'autres éléments. Le résultat se compare à une "valeur ajoutée" qui, au travers de procédés informatiques dont le rôle n'est certes pas négligeable, a comme fonction essentielle de transmettre l'information des spécialistes de disciplines différentes tant à la communauté scientique qu'aux utilisateurs directs - aménageurs ou professionnels du monde agricole.

CARTOGRAPHIC AND ATTRIBUTE DATA BASE CREATION FOR PLANNING ANALYSIS THROUGH GBF/DIME AND CENSUS DATA PROCESSING

Apollo T. Teng
Intergraph Corporation
One Madison Industrial Park
Huntsville, AL 35807

ABSTRACT

The highly topological structure of the U.S. Bureau of the Census GBF/DIME-File provides a sound base for generating cartographic and attribute data bases for planning analyses. The methodology and procedures for implementing GBF/DIME maintenance and a DIME-based planning information system in an interactive computer graphics and attribute data base system are discussed. The dynamic linkages between the graphic elements and the attribute data base allow the updating of the street segment attributes at the same time the graphic element is modified. The system also allows a user to extract population summary statistics from census tapes and store it in the attribute files corresponding to the various graphics files. Additional local data can be geocoded, tallied, and loaded to the attribute data base. The attribute data base can be queried, manipulated, reported, and displayed interactively. The system provides the capability to display polygon thematic data by using crosshatches or polygon color fill. The advantages of the system can be summarized as: responsiveness, immediateness, and ease of operation.

INTRODUCTION

Since 1970 the United States Bureau of the Census has conducted the decenial census by mail. The GBF/DIME system was developed to automate the processing of the questionnaires returned by each household.

The GBF/DIME System

Geographic Base Files are created using the Dual Independent Map Encoding technique which was developed by the United States Bureau of the Census to encode city maps into computer-readable form. The basic unit of record is a street (or nonstreet) segment. A segment is a length of the street between two intersections. There are essentially two sets of information recorded for each segment: one is related to the segment--street name, intersection numbers at both ends of the segment, coordinate values of the two intersections, and a record sequence number; the other is related to either side of the segment--address range, zip code, census block, census tract, place (e.g., city) code, country code, SMSA (standard metropolitan statistic area) code, congressional district, etc.

The Functions of the GBF/DIME System

The original purpose of developing GBF/DIME is to assign geographic codes--such as census block, census tract--to the census questionnaires returned by each household covered in the mail out/mail back census. The geocoded census data is aggregated into various geographic units for reporting purposes. To use the GBF/DIME system in the planning activities of local government, four areas of application can be identified.

Geocoding through address-matching. The capability can be used in any applications which require the assignment of geographic codes to a set of records which have street address as one of its attributes. In a local government environment, examples of application include geocoding welfare cases into census tracts for reporting to the overseeing agency; assigning traffic analysis zone to motor vehicle registrations for trip generation study. In areas outside government, examples of application include geocoding bank customer addresses to census tracts for reporting to the state banking department; geocoding socioeconomic data of area residents into market areas for studying purchasing power, market penetration potential, etc.

Computer mapping. The coordinate values recorded for the two intersections (nodes) of each street segment in the GBF/DIME create a computer map. A plot of all the segments is basically a single line street map. Data related to the street segments or intersections can be displayed through computer graphics techniques. The GBF/DIME can be restructured to provide the vertices for each census block, block group, or census tract contained in the file. The proper chaining of the vertices for each census block, block group, or tract results in a definition of the polygon for graphic display. Coupled with attributes at the tract, block group, or block level, such polygon graphic files provide the basis for thematic data display and analysis.

District delineation. With the census blocks as the basic building blocks and the demographic data associated with these census blocks, a capability can be developed to automatically generate compact districts such as municipal service districts or election districts. This kind of application falls in the general category of districting.

Network analysis. The street network represented in a GBF/DIME can be used in transportation related studies. These include city bus and school bus routing, emergency vehicle dispatching, refuse collection routing, trip generation, mode assignment, and traffic control planning, etc.

CARTOGRAPHIC AND ATTRIBUTE DATA BASE CREATION

The software product for creating cartographic and attribute data bases consists of the following components:

1. A procedure transfers and converts a GBF/DIME tape into interactive graphics and attribute data base environment. The attribute values of the segment entity in the data base are extracted from the attributes associated with each street segment in the GBF/DIME file.

2. An online GBF/DIME edit capability enables an operator to interactively edit the graphic elements (i.e. street segments) as well as review and modify the attributes associated with the edited graphic elements.

3. A program procedure constructs Census Block, Block Group, and Census Tract polygons from the basic graphic elements (lines). Procedures such as area calculation, attribute density calculation, centroid calculation, and polygon attribute display are then possible.

4. A procedure extracts Public Law 94-171 summary population statistics to be loaded to attribute files associated with the Census Block, Block Group, and Census Tract graphic files. Attribute reporting and thematic mapping are then possible.

5. A geocoding procedure for assigning geographic codes to the user data files which are collected at the individual addresses and a data aggregation procedure for tallying the geocoded data into summary statistics for loading into attribute data bases associated with the various polygon graphic files.

The graphic manipulation and hard copy plotting are supported by the nucleus interactive graphics software. Similarly, the creation, updating, and reporting of the attribute data bases are the functions of the nucleus data base management software.

GBF/DIME-FILE PROCESSING

There are two primary reasons for loading a GBF/DIME file into an interactive computer graphics system: (1) To edit, modify, or update the street segments and the attributes of the street segments for the maintenance of the GBF/DIME system; (2) To merge the geographic data which can be derived from a GBF/DIME file with other sources of information in a geographic data base.

GBF/DIME Tape Interface

This program loads the GBF/DIME tape in the standard U.S. Bureau of the Census format into an interactive graphics and data base environment. A graphic element (i.e. a line) with an attribute linkage representing each street segment in the GBF/DIME file is created in the graphics file. A text node is placed at the midpoint of the segment. Attributes can be selected and displayed at the text node. However, the street name is placed only once within each census tract.

In addition, a corresponding attribute insert command line is written to an ASCII file, which can later be used to load all the attribute values recorded in the DIME file to the attribute data base file. For the GBF/DIME maintenance, we have to keep the street segment attribute entities in a separate, tightly controlled data base containing all the relevant attribute values for the street segments. The attribute insert command file for inserting attribute values into the street segment attribute data base is fixed in its format to facilitate the development of controlled editing procedures.

However, a user can modify his data base schema to include additional attribute files for recording attributes such as:
(1) Segment-related: pavement width, right of way, number of vehicular lanes, traffic direction, etc; and
(2) Geographic codes on both sides of the segment: traffic analysis zone, school district, councilmanic district, health service district, tax district, etc.

Interactive DIME File Editing

A series of MACRO command files were developed to assist an operator in editing the GBF/DIME graphics (street segments) file. Most of these user commands also allow for the editing of the attributes associated with the street segments.

Specific interactive GBF/DIME editing functions. The functions (in the form of user commands) provided for on-line editing and maintenance of GBF/DIME segment graphics and attribute files include: (1) Allows digitizing of new street segments, copying of attributes from existing segments, adding new attribute values, adding X & Y coordinate values for the from- and to-node of the segment, and changing of any existing attribute values; (2) Allows modification of existing street segments in the graphics file and automatically changes X & Y coordinate attributes to correspond to new X & Y values of the modified segment; (3) Allows moving of existing street segments in the graphics file, and automatically changes X & Y coordinate attributes to correspond to new X & Y values of the moved segment; (4) Allows copying of attributes from an existing street segment, adding new attributes values, adding X & Y coordinate values, and changing of any existing attribute values; (5) Allows digitizing of new street segments without any data base communication. The user can later use feature 4 to attach and add attributes to the newly digitized street segments; and (6) Guides the user through the digitizer set-up procedures. This is for digitizing a new subdivision or an area with new layout of streets. The source map for the new subdivision is most likely of different scale than the DIME street design file.

Generating an updated GBF/DIME from the data base. A data base formatted report (without headings and page breaks) will generate a GBF/DIME file in the standard Census Bureau format. This file is for data transfer to a noninteractive graphics/computing system.

CENSUS POLYGON CREATION AND ATTRIBUTE LOADING

A GBF/DIME clearly records the various geographic codes on both sides of each street segment. The geographic codes include county code, place code, ZIP code, census tract number, census block group number, and census block number. Those segments serving as the boundaries for each level of geographic areas can be ascertained and written to separate work files which serve as polygon definition files for the various levels of geographic areas.

Constructing Census Polygons with GBF/DIME as Input

These segment records are first split to single-sided records with each record describing a census block side. (Such a file is nicknamed "Nickle File" among the geoprocessing community.) When sorted by the appropriate geographic code(s), these records are grouped together to form the encompassing sides of the polygons. Each group of these records which describe a geographic area is then topologically chained around the polygon and the corresponding vertex coordinates are collected to a polygon vertex chain file, which is later used for creating the polygon graphic files. This procedure applies to all the geographic areas recorded in the GBF/DIME.

Indentation of block vertices and double-line streets. To more closely represent the city geography, it is highly desirable to indent the block vertices so that the street right-of-ways will be shown. This is particularly so when one is using large scale map. In the absence of the true street width information, the system inserts three levels of street width codes to the Nickle file. Freeways or parkways are assigned a width code of 4, all other streets 2, and nonstreets 0. The amount of indentation is guided by these width codes.

Graphic elements with attribute linkages. While each census polygon is placed in the design file, the corresponding attribute linkage for the graphic element is also written to the element header. In addition, an insert command line for loading attributes to the corresponding data base file is written to an ASCII file according to the data base language format. This file is then used to bulk load the polygon attributes to the data base. As a minimum, the attribute values that are written to the insert command file are the polygon IDs, the polygon centroid coordinates, and the area of the polygon.

Loading The Population Census Data

Census summary tapes distributed by the Bureau of the Census are an important source for socioeconomic data.

The summary is done at the various census geographic levels, e.g., census tract, block group, and census block. A mechanism is needed to extract these summary data and load to the respective data base files.

Public Law 94-171 population census summary tape provides data on total population, white population, black population, Asian population, and Spanish surnamed population. The data were aggregated into county, minor civil division, census tract, and census block levels. A software module is provided to extract the data items from the PL94-171 tape and write to three separate work files, i.e., census tract, block group, and block. The processing is at the county-by-county basis. A separate software module matches the geographic codes between the work file and its corresponding data base insert command file. The latter is appended with the population attributes extracted from the PL94-171 tape. A newly calculated attribute value of population density is also appended to the insert command file. These expanded attribute insert command files are then used to load the attribute values to the polygon attribute data bases.

Geocoding User Data

User data which contains street address as one of its attributes can be address-matched against the complete address range inventory in the GBF/DIME street segment attribute data base. Geographic codes such as census tract and block numbers can be appended (geocoded) to the user data records when matches occur. The geocoded data file can then be sorted and statistics tallied at the desired geographic level. These summary statistics can in turn be loaded onto the planning information data base at the various geographic levels.

A software program has been developed for address matching and geocoding of each data record recorded at street address to the approximate location along the street segment as well as the census tract and block that the record falls in. The approximate location was derived from the assumption that the X, Y values of the street address is proportional to the address range of the street segment. A graphic display of these geocoded locations is possible. A separate data aggregation function tallies geocoded data onto various geographic levels and loads the data onto the corresponding attribute entities.

ATTRIBUTE REPORTING AND THEMATIC MAPPING

The attribute file in the data base goes hand-in-hand with the graphic elements in the graphics file. From a graphic workstation, one can interactively query the attribute data base and manipulate the attribute. One can also select and highlight the polygons based on the attribute values. This is exactly the power of an interactive graphics and data base system that supports planning analysis and decision making process.

Annotating Graphics with Attributes

To conserve disk space and to improve legibility, the street names are placed only once for each census tract. However, when the need arises to annotate the street segments with some of the attributes stored in the DIME data base, the displayable attribute processor can be utilized to annotate the segment graphics file with the selected attributes. The annotation is placed relative to the location of the text node placed along with the street segment.

This capability enables one to generate almost unlimited cartographic products. The wealth of the attribute data base can be tapped for annotating the basic line art for different purposes.

Attributes Reporting

Attributes can be selected and reported through the data base language. If executed from a graphic station, the attribute selection can be further guided by the graphic manipulation commands such as layer selection, windowing, or fence.

Thematic Mapping

To graphically represent the attribute values associated with the various levels of census areas, either the area color fill or crosshatching technique can be employed. A preprocessor would retrieve the attribute values for the polygons to be crosshatched from the polygon attribute data base. This attribute file along with the crosshatching parameter table will be input to the crosshatching module for the actual patterning of the polygons. The parameter table defines the crosshatching angle, line weight, line style, color code, crosshatching density for each value range.

CONCLUSION

The 280 GBF/DIME-Files that the U.S. Bureau of the Census created for the same amount of Standard Metropolitan Statistical Areas provide a wealth of cartographic data base for local government personnel in monitoring and displaying the aggregate physical, social, and ecomic conditions of the communities at the various geographic levels. The interactive computer graphics technology makes the maintenance of the GBF/DIME-Files and the creation of the various cartographic files possible. Furthermore, the associative attribute data base capability enables the wealth of census summary data to be tapped at the fingertips of the analysts. The geo-coding function further enriches the planning data base with the various locally collected data. The efficient creation, manipulation, and presentation of the graphic and attribute data bases give a new horizon for the planners in their analysis, planning, and decision making processes.

COMPUTER MAPPING FOR BIOMASS INVENTORIES

W.C. Moore and W. Chow
Remote Sensing Project Leader and
Regional Computer Systems Manager
Northern Forest Research Centre
Canadian Forestry Service
Environment Canada
5320-122 Street
Edmonton, Alberta T6H 3S5

ABSTRACT

The Northern Forest Research Centre is developing a computer mapping capability as part of a geographic information system for the western and northern region. The initial requirement was to develop a capability for converting forest inventory maps and supporting data to a biomass data base. The location-specific data manipulations for a test inventory map from Saskatchewan are described and evaluated. This is more than just a technical challenge, however, because while experts make maps, nonexperts use them to make policy. The cost-effectiveness of automated reinterpretation of existing data is discussed, as is the adequacy of such reinterpretation. Future requirements for remote sensing and other integrated inputs are also considered. Resource analysts and policy makers require that quantitative storage, classification, manipulation, and retrieval of location-specific results (as maps and tables) are reliable and available on demand. A complete resource analysis, mapping, and information system must be designed and developed for practical and continuing operations in a clear, timely, accurate, and economic manner. Further developments and operational applications are required.

INTRODUCTION

Combustion of forest biomass, a term that includes all forest vegetation, already accounts for about 4% of Canada's energy needs. The forests have great promise as a renewable energy source. National considerations, therefore, should immediately involve inventories of how much forest biomass is available and where it is located. This paper examines how the large quantities of existing forestry maps and supporting data can be transformed into reasonable biomass estimate inventories.

The development and operational applications of computer mapping systems are important bases for geographic information systems. Biomass inventories, or data bases, can become an integral part of such a system. A computer mapping system is basically an efficient map filing system whereby spatially referenced data can be efficiently stored, classified, updated, manipulated, and retrieved as custom maps and tables. This is a significant improvement over the traditional banks of map drawers, thematic overlay products, map redrawing, and tedious dot counting now necessary to obtain the required forestry data. As a practical example of computer mapping development, a system has been installed and a sample forest inventory map has been converted to a biomass data base at the Northern Forest Research Centre (NoFRC). This can be used to illustrate some of the practical advantages, potential limitations, and future developments of such a s tem.

The basic components and theoretical operations of the NoFRC mapping system were described previously (Kirby and Chow, 1982). Practical operations have more recently been evaluated using a sample forest inventory map and supporting data (i.e., stand and stock tables) provided by the Forestry Branch of the Saskatchewan Department of Tourism and Renewable Resources. The conversion of this material to a biomass data base was possible through the application of regional biomass equations to forest species volumes.* The renewable biomass of potentially overlapping peatland inventories might be an additional requirement in the future (Zoltai and Pollett, 1983). This would be another reinterpretation of existing data, but it would also be a feasible compromise between budgetry constraints and increasing information demands (Napton and Luther, 1981).

The utility of computerized mapping of summarized forestry statistics covering the whole country has been demonstrated using the considerable facilities of Statistics Canada (Cunningham, 1980). Canadian Forestry Service regional establishments, such as NoFRC, have research, assistance, and coordination responsibilities for more detailed operations and statistics in cooperation with provincial and other forest agencies. Thus, additional work is required to make the NoFRC mapping system compatible with a variety of map and remote sensing inputs that have been documented by others (Myers, 1981, and Brooner, 1981). This is a next step, because computerized systems also increase operational productivity and flexibility, particularly for multisource change mapping (Milazzo, 1981, and Wilson and Thomson, 1981). These developments provide new, practical capabilities. For example, cost-effective means of using Landsat data have already been demonstrated operationally in the revision of topographic maps (Fleming, 1982) and the monitoring of burned forest land in remote areas (Moore, 1983). In addition, forest mensuration with large-scale aerial photography has been developed to significantly enhance fieldwork (Kirby, 1980). Nevertheless, it is important to realize that while experts make maps, nonexperts use them for policy making. Great care is therefore required with manipulations of existing data, and with concise qualifications of the results (Napton and Luther, 1981).

---

* from T. Singh, "Conversion of Tree Volume to Biomass in the Prairie Provinces", a proposed <u>Forest Management Note</u>, currently under review, NoFRC, Edmonton

METHODOLOGY

Mapping System

The complete Mapping and Analysis of Resources System (MARS) is organized according to Figure 1. In that flow chart, the hatched lines enclose the Systemhouse Limited* Resource Analysis and Mapping System (SHL-RAMS) that is central to MARS. Developments to the summer of 1983 have progressed to the biomass computations stage. In addition, the Gregory Geoscience** Procom-2 image transfer and mapping system for the change interpretation stage has been acquired separately. The numbers between the blocks, or stages, in Figure 1 refer to the operations that are briefly described in Table 1. The larger arrows to the right of the flow chart indicate the outputs of MARS. Future growth potential of the system is described later.

Biomass Mapping

The whole 1 : 12 500 scale (i.e., for provincial field operations) Bittern Lake forest inventory mapsheet (UTM Z13 E45 N597) was processed to form a biomass data base. Volumes by tree species and cover type (i.e., dominant, subdominant, and understory) from the stand and stock tables were converted directly to biomass per hectare by forest inventory polygon. The result was a biomass inventory map. A portion of that map (i.e., a 4 km^2 test area) is reproduced, with principal tree species annotations, in Figure 2.

The volume data in the stand and stock tables, however, was for merchantable volumes only. Some assumptions had to be made for unmerchantable stands. In this case, therfore, the biomass estimates might be a little high. Mean stem heights and diameters were also available in the tables, but were not used for data processing and consistency reasons. Appropriate equations also exist for biomass conversions using mean heights and diameters for this region (Singh, 1982) and might be used in supplementary sampling later.

The data for the Bittern Lake mapsheet was manipulated as follows:

1. After digitizing and correcting for map errors, the polygons were sorted by species association, height class, density class, and the three cover types. The polygon identification numbers and areas were stored for future use. Applicable portions of the stand and stock tables were also entered as a volume table.

2. The data was printed and stored on the PDP 11/60 compnter.

3. Volume estimates were calculated using the data in the stand and stock tables. This was done by matching map attribute listings to the volume table variables as follows:

---

* Systemhouse Limited
2827 Riverside Drive
Ottawa, Ontario K1V 0C4

** Gregory Geoscience Limited
1750 Courtwood Crescent
Ottawa, Ontario K2C 1B5

Mapping & Analysis of Resources System (MARS)

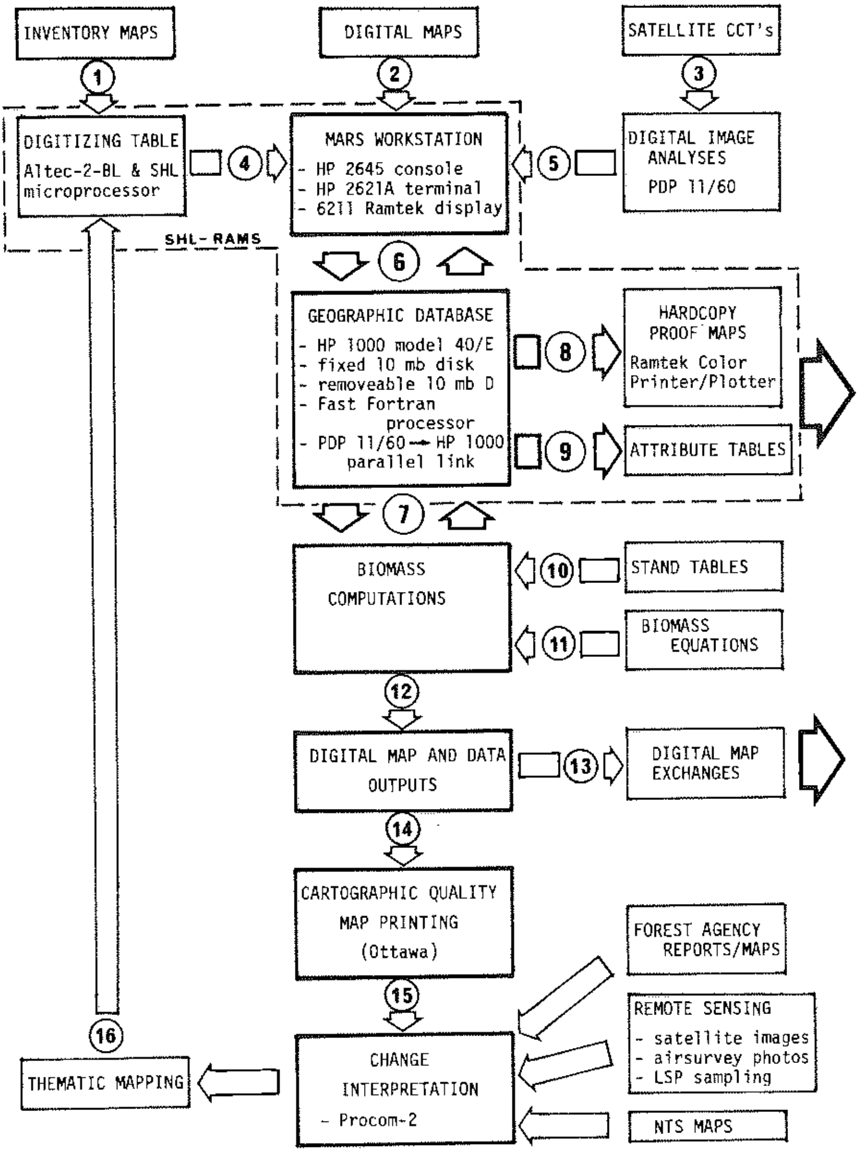


Figure 1. NoFRC Computerized Mapping System
Abbreviations are as follows:
CCT's - computer compatible tapes (i.e., Landsat)
LSP - large-scale photography (aerial)
NTS - National Topographic Series
GIMMS - Geographic Information Manipulation and Mapping System (Cunningham, 1980)

MARS Operations

1	A high-quality (preferably mylar) line map transparency is required, and geographically referenced attributes, or data definition files (DDF), are to be designed.
2	OPEN map file - digital map ready for display on screen.
3	Earth satellite image data is received as CCT's by user.
4	SHL-RAMS commands on digitizer to enter line segments.
5	Unsupervised classifications have been used with parts of CCT image and converted to RAMS-compatible display.
6	RAMS commands to PAINT or DRAW required for Ramteck display. Other commands to display tables or generate reports. Can be transfered to PDP 11/60 for printing.
7	Biomass variables computed and entered as part of DDF.
8	RAMS PRINT command will produce map hard-copy of the Ramtek display on the Ramtek plotter.
9	LEGEND command produces attribute classes on display, and PRINT command will produce a hard-copy reproduction.
10	Stand tables required with digitized maps for sufficient data for biomass prediction equations (e.g., volumes).
11	Biomass/hectare computed by polygon and entered in DDF.
12	Each attribute can be displayed independently or combined with other attributes for plotting thematic maps.
13	RAMS ◄══► GIMMS provided with the system for compatibilities with other computer mapping systems.
14	GIMMS-formated tape can be sent out for production of a cartographic-quality line map.
15	A cartogrpahic-quality map can be a base map for overlaying new map themes to form new polygons. Procom-2 now available for updating maps for burned or cut-over forest lands from Landsat or integrating other data.
16	Existing DDF can be upgraded by digitizing change maps on cartographic-quality map of existing DDF. Entire map would not have to be redigitized because of a new polygon overlay capability of SHL-RAMS.

Table 1. Itemized Operations in the Figure 1 Flow Chart

(i) Where map attributes matched volume table variables, the biomass equations for the species were applied. These were summed to determine the total estimated biomass weight per hectare for that polygon, which was then multiplied by the area of the polygon to tabulate the total biomass estimate for the polygon.

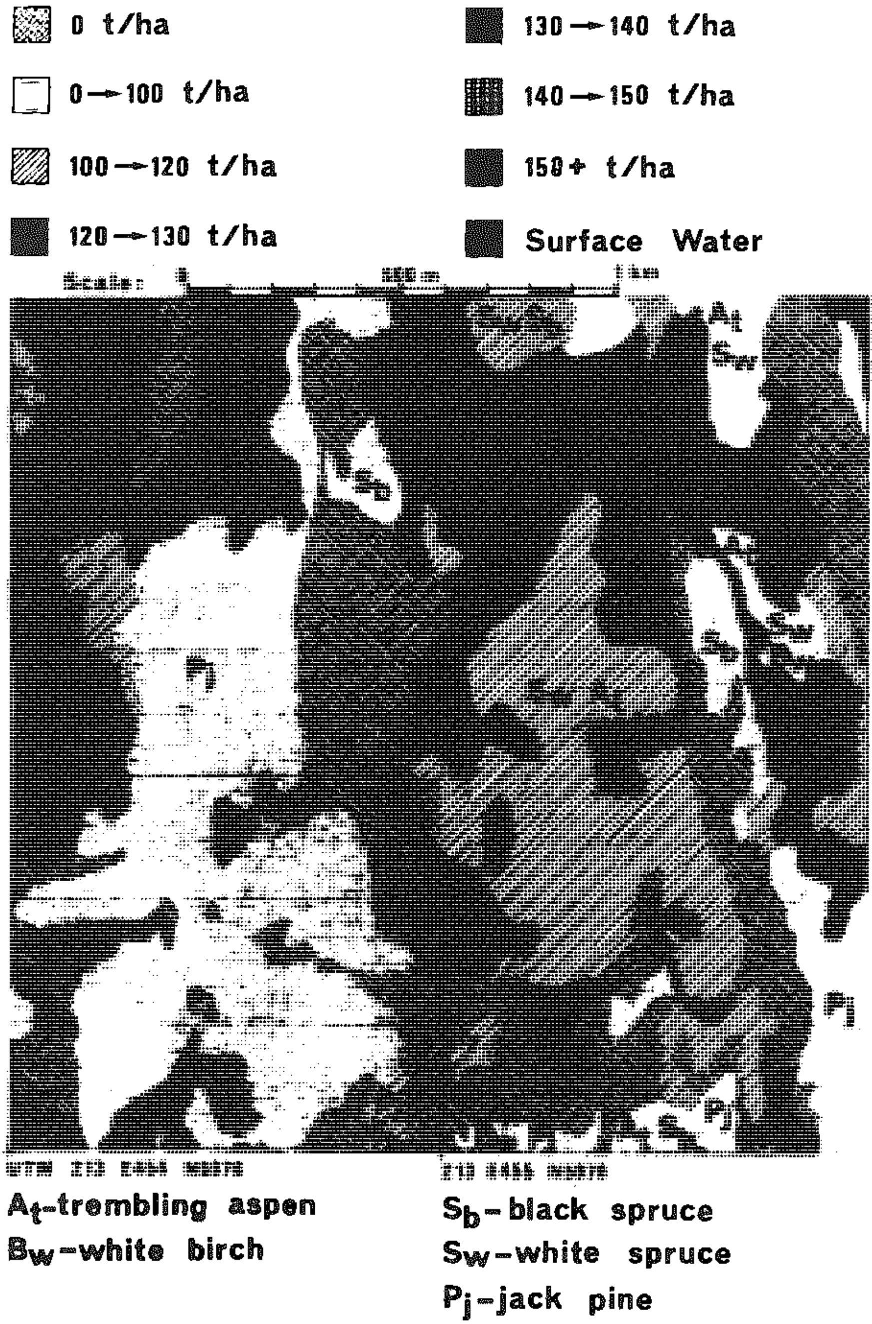


Figure 2. Biomass Inventory Map

(ii) When an exact match was not found, the cover types were reversed and the search was rerun. If there was then a match, the procedure in (i) was followed.

(iii) If there was still no match, the second cover type was deleted and the search was rerun on the first cover type. If there was then a match with the volume table, the procedure in (i) was followed.

(iv) The last resort was to print out the polygon information, compare it to the original map, and make whatever corrections were necessary. When there was still no match, a species was substituted (i.e., one that was similar), and the calculations were done manually.

4. The estimated biomass weight per hectare and total biomass weight per polygon variables were defined and entered in a data definition file (DDF) on SHL-RAMS. This was used to produce Figure 2.

Programs could be written on the NoFRC PDP 11/60 minicomputer to sort and print various selections of location-specific data from the data base, but this was not deemed appropriate. Instead, computing the biomass before the map is digitized appears to be a better idea for the future. The DDF could then be established to accommodate these values, and all the sorting required would be done on SHL-RAMS.

DISCUSSION

The size of the area of forested land in the western and northern region of Canada might at first appear to be overwhelming. Over 10 000 forest inventory maps similar to the Bittern Lake example from Saskatchewan are required to cover the forested land on the prairies. As many again are required for the forested land of the Northwest Territories. Forestry agencies in the provinces of Alberta, Manitoba, and Saskatchewan, however, are placing considerable emphasis on compiling and updating their forest inventories. In addition, these provinces are in various stages of establishing computer mapping facilities of their own. Forest management in the Northwest Territories is at present the responsibility of the federal Deparment of Indian Affairs and Northern Development, which is not nearly as advanced in forest inventory mapping and sampling as are the prairie provinces. There is great potential for more cooperation among all these forestry agencies for assessments of such a huge area, and biomass computations are a natural follow-up to forest inventory surveys.

The forests of the region lie principally in the boreal and transitional forest regions — parklands and foothills to the south and tundra to the north. The forest cover is largely heterogeneous. The significance of the extensive peatlands in the forested areas as an energy source is a separate question entirely, but location-specific peatland classifications might one day be defined and included in a biomass data base. As has been shown, the forest inventory emphasis on merchantable timber can also be a problem for complete biomass computations. Nevertheless, the forest inventory maps and supporting data provide the best information available for the creation of data bases and calculations of biomass estimates. They currently include data on the bulk of the aboveground biomass. There might be further improvement with supplementary sampling in the future. Independent biomass surveys do not appear to be cost-effective for such a complex and extensive forested region.

Remote sensing offers considerable information on the dynamics of such large forested areas (e.g., clear-cutting, burned, insect/disease mortality, flooding). This can provide valuable information for updating forest, and subsequently biomass, inventories — particularly when forest surveys are completed on 10-, 20-, or 30-year cycles. Normal forest inventory sample plot procedures might also be enhanced to include more total biomass measurements to improve estimates. In addition, the national Forest Insect and Disease Survey has a requirement for a multitude of thematic maps that might be included in a computer mapping system. Efficient computer mapping is necessary to integrate the multiple sources of relevant information in a cost-effective manner.

A comparison between a computer mapping system and a word processor might be useful for gaining an appreciation of the technology involved. Word processors operate on the basis of creating and manipulating computer data files. These files contain words, numbers, symbols, sentences, paragraphs and tables that can be used to form a text. This text can then be revised and adjusted quite efficiently, and it can be automatically printed in a variety of selected formats. A computer mapping system goes beyond the well-developed word processor technology application to involve the creation, storage, and retrieval of geometrically referenced data files for points, lines, polygons <u>and</u> texts. In addition, the data files are often further processed within the system to create new, geographically referenced files such as biomass estimates for polygons. Systematic revision, adjustment, or updating of files in the considerably less-developed computer mapping technology is somewhat more complex and time consuming than similar operations in most word processors. Nevertheless, the NoFRC system is useful for experimental and training purposes, and it has the potential for great improvements in operating capabilities at comparatively low cost. It also has the potential of interfacing with provincial systems as they become operational, thereby enhancing developments in remote sensing applications and sampling designs at NoFRC.

The SHL-RAMS portion of MARS might be considered a modular component that can be modified or replaced with little or no effect on other components. RAMS was a prototype development of Systemhouse Limited. Improvements have been made to the system, but it is still too slow for operational requirements. Over 40 hours were required to digitize the approximately 2000 polygons of the example Bittern Lake map. Improvements in the technology need to be implemented, and several options are being considered to update the SHL-RAMS.

CONCLUSIONS

1. Computer mapping has an immediate potential for functionally upgrading geographic information systems. The technology is evolving at a high rate, and such systems should approach the efficiency and utility of modern word processors in the near future. In practical terms, these developments should provide the required productivity increases to meet the increasing environmental information requirements within constrained budgets.

2. Data definition files in the current NoFRC SHL-RAMS should be defined and established before maps are digitized. Technology improvements are also required to ease the current bottleneck in digitizing.

3. Caution is required in manipulating existing location-specific classifications and supporting data to produce maps and tables for which the original field survey was not necessarily intended — particularly for a very large area that is as heterogeneous and dynamic as the boreal forest region.

4. Supplementary sampling for biomass (i.e., all forest vegetation) might be required for more complete tabulations. In practical terms, however, quantitative biomass samples pertaining to lesser vegetation and regeneration areas might have to be systematically included with normal forest plot measurements to be economically feasible. Overlapping peatland sampling should be another consideration as an additional renewable biomass source. Various types of remote sensing applications might also be developed to provide supplementary biomass estimates in a location-specific geographic format.

5. Functional digital map exchanges and remote sensing integration for monitoring significant changes to the ground cover have yet to be tested with the NoFRC mapping system.

REFERENCES

Brooner, W.G. 1981, An Overview of Remote Sensing Input to Geographic Information Systems: Remote Sensing: An Input to Geographic Information Systems in the 1980's; Proceedings of the Pecora VII Symposium, B.F. Richason, Jr. (ed.), pp. 318-329, American Society of Photogrammetry, Falls Church

Cunningham, R.A. 1980(?), Computerized Mapping of the National Forest Inventory, Geocartographics Subdivision, Special EDP Services Division, Statistics Canada, Ottawa

Fleming, E.A. 1982, Topographic Map Revision Using Satellite Imagery: Proceedings of the 2nd National Workshop on Engineering Applications of Remote Sensing, pp. 149-159, Canada Centre for Remote Sensing, Ottawa

Kirby, C.L. 1980, A Camera and Interpretation System for Assessment of Forest Regeneration, Information Report

NOR-X-221, Northern Forest Research Centre, Canadian Forestry Service, Environment Canada, Edmonton

Kirby, C.L. and Chow, W. 1982, A Mapping and Resource Analysis System (MARS) at the Northern Forest Research Centre. Proceedings of the Canadian Institute of Surveying; Centennial Convention, pp. 367-386, Canadian Institute of Surveying, Ottawa

Milazzo, V.A. 1981, The Role of Change Data in a Land Use and Land Cover Map Updating Program: Remote Sensing: An Input to Geographic Information Systems in the 1980's; Proceedings of the Pecora VII Symposium, B.F. Richason, Jr. (ed.), pp. 189-209, American Society of Photogrammetry, Falls Church

Moore, W.C. 1983, Operational Mapping of All Burned Forest Land in the Northwest Territories with Satellite Imagery, Forestry Statistics and Systems Branch, Canadian Forestry Service, Environment Canada, Chalk River (in press)

Myers, W.L. 1981, Integration of Remotely Sensed Data into Geographic Information Systems; Workshop Session Outline: Remote Sensing: An Input to Geographic Information Systems in the 1980's; Proceedings of the Pecora VII Symposium, B.F. Richason, Jr. (ed.), pp. 3-14, American Society of Photogrammetry, Falls Church.

Napton, D.E., and Luther, J. 1981, Transferring Resource Interpretations: Limitations and Safeguards: Remote Sensing: An Input to Geographic Information Systems in the 1980's; Proceedings of the Pecora VII Symposium, B. F. Richason, Jr. (ed.), pp. 175-186, American Society of Photogrammetry, Falls Church

Singh, T. 1982, Biomass Equations for Ten Major Tree Species of the Prairie Provinces, Information Report NOR-X-242, Northern Forest Research Centre, Canadian Forestry Service, Environment Canada, Edmonton

Wilson, C.L., and Thomson, F.J. 1981, Integration and Manipulation of Remotely Sensed and Other Data in Geographic Information Systems: Remote Sensing: An Input to Geographic Information Systems in the 1980's; Proceedings of the Pecora VII Symposium, B.F. Richason, Jr. (ed.), pp. 303-317, American Society of Photogrammetry, Falls Church

Zoltai, S.C., and Pollett, F.C. 1983, Wetlands in Canada: Their Classification, Distribution, and Use, Chapter 8: Mires: Swamp, Bog, Fen and Moor, B. Regional Studies, A.J.P. Gore (ed.), pp. 245-268, Elsevier Scientific Publishing Company, Amsterdam

FLOOD ESTIMATION IN EUROPE - A CASE STUDY OF APPLIED DIGITAL CARTOGRAPHY

S.E. Wiltshire
Institute of Hydrology
Wallingford, Oxfordshire OX10 8BB
England

ABSTRACT

A European flood study has prompted the development of a system for automatically deriving various catchment-based indices of climate and physiography. Primary thematic mapping is digitised by a variety of techniques and is translated to values on a 2.5 km data raster. Catchment boundaries are also digitised and their contained areas stored in a raster form so they can be efficiently associated with the thematic data. The use of rastered data, together with a flexible processing system, provides a powerful tool in the development of flood prediction procedures.

INTRODUCTION

The European Community has commissioned a European flood study, which has as its object the provision of flood estimation procedures for the territories of member-states. Information on the flood flow of rivers is available at gauging station sites, where a historical record of peak flows can be used to specify the frequency with which a particular flood flow is exceeded. For flood estimation at sites without recorded data it is necessary to transfer available hydrological information from gauged sites. One transfer technique is to use multiple regression analysis to relate flood flow statistics empirically to climatic and physiographic characteristics of catchment areas obtained from maps.

The traditional method of deriving a catchment characteristic is to overlay the catchment boundary on the thematic map of interest, and by square counting, planimetering or otherwise, determine the average value of the mapped quantity. The European Flood Study incorporates over 1800 catchments and seven mapped characteristics. The scale of the project is such that a revision of methods used to derive catchment characteristics was found to be necessary.

It was decided to automate the procedure as far as possible, and a processing system has accordingly been developed that can manipulate digitised catchment boundaries together with rastered thematic data, so that the overlay process can be performed automatically. This approach has demanded a large investment in program development and digitising effort, but this is justified by the ability to derive easily a wide variety of catchment characteristics and the facility to replay digitised or derived maps.

Reduction of data to a gridded or raster form varies with different types of variable. For continuous variables, such as average rainfall, this is performed by the SURFACE 2 gridding and contouring package (Sampson, 1975), which is implemented on a Honeywell mainframe computer. SURFACE 2 allows close user-control of search techniques, interpolation method, grid size and error analysis and is a key component of the processing system. The availability of such a package was an important factor in the decision to develop a raster database. Other factors leading to the choice of a raster system in preference to a vector system include the relative simplicity of data storage structures and the ease with which subsets of rastered data can be accessed and manipulated.

The following sections describe experience with various digitising methods, the reduction of digitised data to a raster and the numerical estimation of catchment characteristics.

RASTERED DATA AND THE UNIVERSAL TRANSVERSE MERCATOR CO-ORDINATE SYSTEM

A principal requirement of the European Flood Study processing system is the ability to marshal and reference easily large volumes of data covering extensive geographical areas. As national map series in the study area are published in varying map projections, and as the format of many thematic maps is unspecified (eg. a 'bare' dyeline of rainfall isohyets), the need for a unifying reference system of rectangular co-ordinates was soon recognised. The Universal Transverse Mecator (UTM) co-ordinate system was chosen for this purpose, because it provides

a) unique referencing in metres in a plane rectangular system of co-ordinates;

b) 'continuity' over sizeable areas or zones (ie. 6^{o} of longitude) coupled with a minimum number of zones;

c) minimal scale distortion within a zone (less than 0.04%);

d) transformation formulae from one zone to another and to and from geographical co-ordinates.

In addition, two international topographic map series, the 1404 and 1501 series, are available with a UTM grid, making digitising and referencing tasks relatively straightforward. These maps, at 1:500K and 1:250K respectively, are suitably scaled and marked for many of the basic tasks of the study.

Thus the rectangular UTM grid system is an ideal framework for the rastered data sets. These can be visualised as sets of three rectangular grids corresponding to each of the three UTM zones in the study area. These zone rasters are notionally expanded in the east-west direction to represent bands of land 800 km wide, which allows each catchment to be associated exclusively with one expanded zone. UTM grids of adjacent zones are inclined relative to one another, so approximations are sometimes necessary when transferring grid cell values from one zone raster to the cells of an expanded neighbouring raster.

The cell size of these rasters has been fixed at 2.5 km. Cell position can be stated in UTM co-ordinates and cell corners can be located at integer multiples of the cell width. The choice of cell size was made with particular regard for the accuracy of calculations performed on subsets of the rastered data. A rule-of-thumb requirement of computational accuracy is for each catchment to contain not less than nine cell nodes. Catchment areas vary between absolute limits of 0.1 to 100,000 km^2 with a modal range of 100-500 km^2, so a cell unit of 6.25 km^2 equivalent area meets this requirement for the majority of catchments. Sub-cells are defined by interpolation where necessary for smaller catchments. Other factors bearing upon the question of grid size include the scale of primary or source mapping, the scale of map plotting and the limitations of computer storage.

*Conversion from digitised or 'table' co-ordinates to UTM co-ordinates is effected by 'rubber sheeting' (Monomonier, 1982), which requires fiducial points to be defined on the map of known position in both co-ordinate systems. This empirical adjustment is preferred to a process based on explicit consideration of map projection, because it accommodates rotation scale change, projection change and paper stretch with an acceptable positional error. This error increases as the map scale decreases but does not exceed 1km on the ground for a digitising scale of 1:1m.

In-house digitising hardware consists of two D-MAC electromechanical tables and one solid state digitising tablet linked to a microcomputer. Methods of digitising and assigning raster-cell values reflect to a degree the available hardware. These techniques are described for a range of data types in the following section.

CATCHMENT BOUNDARY DIGITISING AND RASTER-REPRESENTATION

Catchment boundaries are often nested or adjacent to one another so that boundary segments are shared by two or more catchments, as shown in figure 1a. A program has been devised to allow for 'borrowing' of segments from other catchments. Figure 1a shows such a situation: Boundary 17 is only digitised along the route BGD; segments BC and CD are borrowed from boundaries 21 and 22 respectively. A boundary string is thus reconstructed to form a continuous loop as shown in Figure 1b; it is then stored in vector form to facilitate subsequent plotting. The effort required to develop 'borrowing' software has been more than offset by savings in digitising time and by the ability to 'play back' adjacent catchments with a single, 'clean', common boundary.

Nodes on a thematic raster contained by a boundary loop could be identified by point-in-polygon analysis, but repeated applications would be demanding of computer time, so the contained areas of boundaries are themselves represented in a raster form. This is achieved by first fitting a minimum containing rectangle of 2.5 km cells to the boundary. Cell-centres lying within the boundary are identified by point-in-polygon analysis and assigned a value of one. All cells outside the boundary are given a value of zero, as shown in figure 1c. The contained area can thus be represented as a matrix of ones and zeros. In practice this must be accompanied by:

1) the boundary identifier (the gauging station number);

2) the co-ordinates of the lower-left corner of the containing rectangle;

3) the dimensions of the rectangle;

4) the grid interval (usually 2.5 km).

Run length encoding is used to reduce further the volume of held data, resulting in a very compact form of the boundary area that can be readily associated with the thematic rasters.

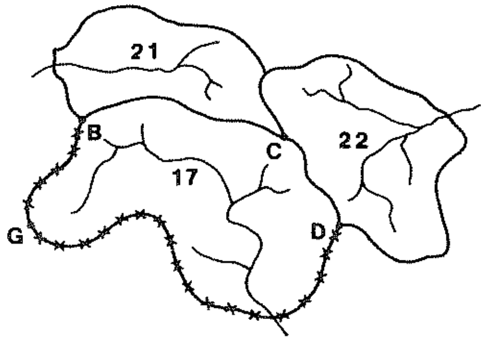


Figure 1a Digitising Catchment Boundaries. Boundary 17 is digitised along the segment BGD, the segments BC and CD are 'borrowed' from boundaries 21 and 22 respectively.

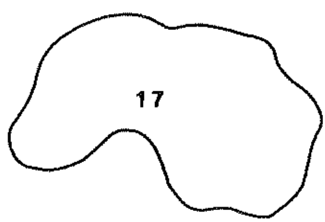


Figure 1b A complete boundary loop reconstruction from digitised and borrowed segments.

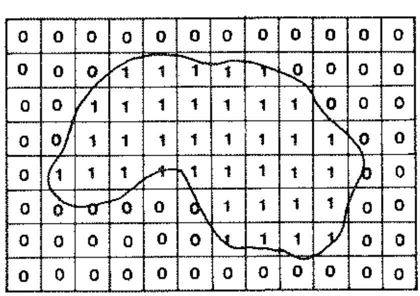


Figure 1c Conversion of the boundary polygon to a sequence of 1's and 0's.

THEMATIC DATA TYPES AND DIGITISING TECHNIQUES

Thematic data employed by the European Flood Study can be grouped into three distinct classes (Beran, 1982):

1) continuously varying quantities such as mean annual rainfall or altitude, usually described by isolines;

2) data of a cardinal nature such as soil type or forest, which appear on maps as a patchwork of constant values;

3) variables defined only along the river network, such as drainage patterns or stream slopes.

Various digitising strategies have been developed to accommodate these data types and these are described below by reference to specific examples.

Rainfall Maps (Type 1)

No single rainfall map is available for the entire study area at a sufficiently large scale. Maps of average annual rainfall have consequently been culled from atlases, yearbooks and copies of unpublished material, and have been redrawn at a common scale of 1:1 M on an orohydrographic version of the Operational Navigation Chart (ONC) series.

Digitising is straightforward: the isopleth value is entered from the keyboard and the line segments digitised from start to end. No provision is made for reconnecting isopleths that extend over several maps, as a good reproduction of the isopleth pattern can be achieved by threading contours through the gridded data, as shown in figure 2. This form of replay is also the basis for routine validation of digitised type 1 data.

Forests, Towns and Lakes (Type 2)

Source mapping for land cover data is the 1:250K, 1501 topographic map series, which presents such information with an acceptable degree of generalisation and consistency. Polygons of forests, urban areas and lakes are too numerous and too intricate to digitise individually and would produce a level of detail not required by the study. A quicker, more approximate method is therefore employed that uses a 1.25 km (5mm) grid as the basic recording unit. This second grid is overlaid on the 1501 map and aligned so that these 1.25 km squares form quadrants of the basic 2.5 km grid cells. The cursor is then moved systematically across the 1.25 km grid and the centre of each quadrant (containing more than 50%) of, say, forest, is logged. The processing program then assigns ones and zeros to logged and unlogged quadrants, and builds up the bit pattern in each 2.5 km cell, as shown in figure 3. A finer grid is used for forests, towns and lakes since it is intrinsically more variable than other data types, and is likely to be of specific relevance to small catchment studies.

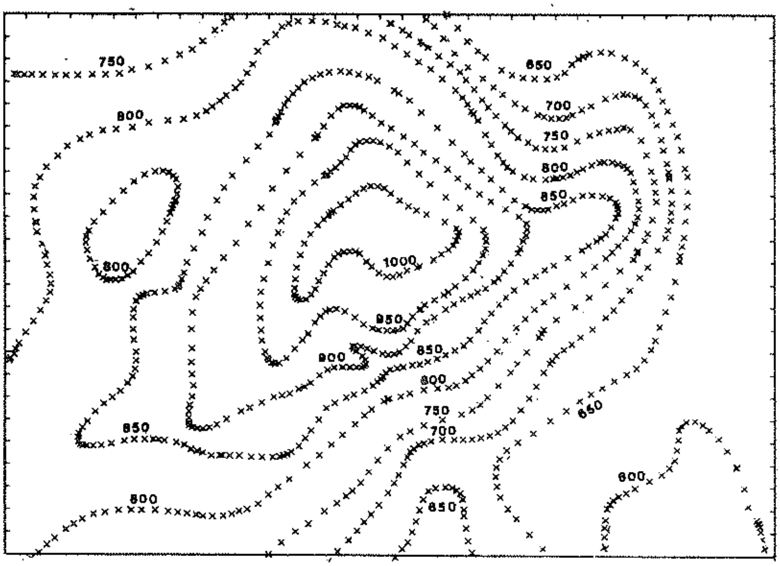
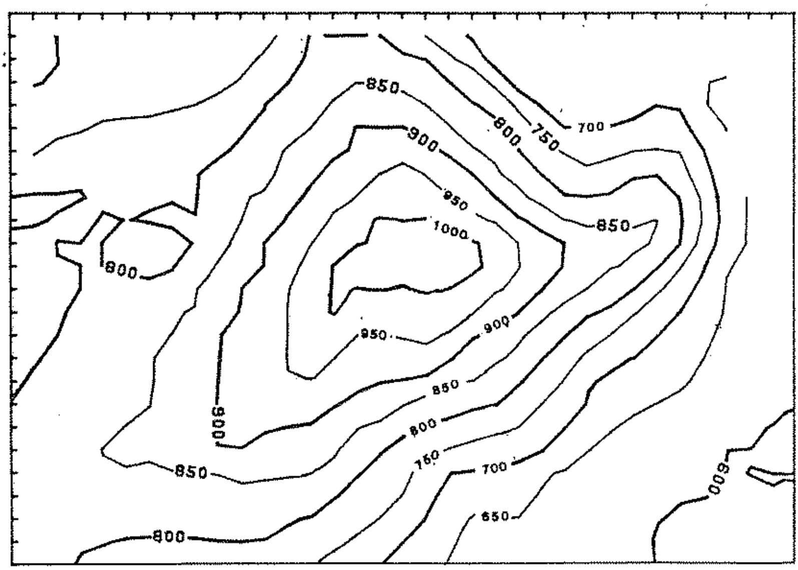


Figure 2 Digitised rainfall isolines (above) compared with those produced by contouring the gridded rainfall data (below). This contour plot has been produced from a grid of 816 nodes.



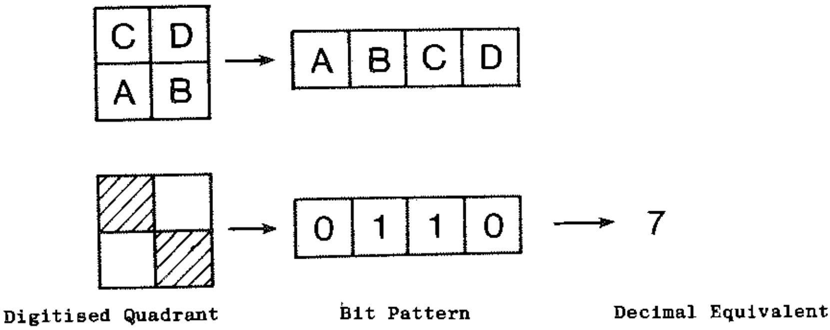


Figure 3 Reduction of digitised forest, town and lake data to representative integers

Since these 2.5 km cells correspond to the raster cells, values can be copied directly to the zone rasters. Some interpolation is necessary to complete the overlap region of the zone rasters.

The simple reduction of digitised data to a binary form allows a highly efficient packing of the raster data; a space-saving of a factor of 14 is possible on a 32 or 36 bit machine.

Soil Maps (Type 2)

A set of soil maps have been specially prepared for the study at a scale of 1:1M by agricultural and hydrological services within the EEC. These maps show five hydrological soil types in a continuous mosaic of polygons. The preparation of these maps and their hydrological application is described by Gustard (1983).

Soil Polygons are less numerous and more generalised than those of forests, towns and lakes, so that digitising polygon boundaries is practical. This is done (one soil type at a time) using the tablet digitiser, coupled to a BBC (colour graphics, 64K) microcomputer and monitor. A transformation is applied to scale the screen display to a UTM-orientated rectangle, and soil polygons are displayed as they are digitised. Software is then used that fills all polygons of a given soil type. Since a screen image consists of strings of 'lit' pixels, an implicit vector to raster conversion is achieved by system hardware. Further, since on the BBC micro there is a direct correspondence between screen pixels and memory locations, a completed image of the rastered version of one soil type can be read from screen memory and transferred to the mainframe zone soil rasters. This procedure avoids the need for point-in-polygon analysis as required for catchment area processing.

Drainage Density (Type 3)

The usefulness of indices derived from the mapped stream network is lessened somewhat by the fact that the representation of blue line information tends to owe as much to a desire for cartographic elegance as to an orderly and consistent reduction of surveyed or remotely sensed data. However, channel network indices have been found to be significant in other flood studies and so a suitable index has been defined and derived for the European project. Since the task of coding and and digitising the complete stream network at a medium scale is too large to contemplate, an alternative approach based on areal sampling has been adopted. This procedure is as follows:

i) overlaying a circle of standard diameter on each 1:50K topographic map in turn;
ii) count the number of intersections between the circle and the mapped channel network and note the co-ordinates of the circle centre;
iii) use SURFACE2 to regrid the 2500 data points associated with map centres to a new 2.5K grid that is registered with the thematic data rasters.

This variable can be remapped in isoline form for inspection or presentation. Thus a type 3 variable may be converted to a type 1 variable for operational convenience.

Stream slopes (type 3)
The slope of the longest stream in each catchment is derived by digitising tbe stream's position and coding the location of each contour crossing. Thus the long profile of the river can be reconstructed and a number of slope statistics derived.

Streamslopes data is by definition catchment-specific and so is not stored on the raster; derived statistics may be written directly to a master library of catchment characteristic values.

APPLICATIONS

The principal use of the gridded data sets is the calculation of catchment characteristics. The manual method of obtaining a catchment-average value of a characteristic is outlined in the introduction. The automated equivalent is now described, using the calculation of a soil index as an example.

CATCHMENT RASTER (A)

0	1	1	0	0
0	1	1	1	1
0	1	1	1	0
1	1	1	0	0
0	1	0	0	0

SOIL RASTER (B)

2	2	3	4	5
3	3	2	2	4
3	2	1	2	2
1	1	2	3	2
2	3	4	5	3

COMBINED RASTER (C)

0	2	3	0	0
0	3	2	2	4
0	2	1	2	0
1	1	2	0	0
0	3	0	0	0

Catchment-Average Soil Type = (2 + 3 + 3 + 2 + 2 + 4 + 2 + 1 + 2 + 1 + 1 + 2 + 3)/13 = 2.15

Figure 4. Derivation of a Catchment-Average Soil Type.

The details of the location and dimensions of a containing rectangle of a catchment can be used to pull from the appropriate soil zone-raster the soil type (defined as either 1, 2, 3, 4 or 5) of each grid cell within the rectangle. This is represented by box B in figure 4. Box A shows the catchment boundary in its grid form. If the elements of B are multiplied by the elements of A, the resulting list of non-zero values is the list of soil cell-values contained by the catchment (box C). It is then a simple matter to obtain the average value for the catchment.

In addition to calculating rapidly average catchment values it becomes possible to determine any other desired statistic of the contained cell values, such as their variance or skewness. Also, with a knowledge of the gauging station position, the distribution of a particular characteristic within a catchment can be examined. These hitherto unmeasured properties constitute a new generation of catchment characteristics that may be useful in developing more accurate flood estimation procedures.

SUMMARY AND CONCLUSIONS

This paper has described the use of digital cartographic techniques in the derivation of catchment characteristics for use in a European flood study. A variety of methods of transferring mapped information to a data raster have been described which accommodate the differing types of thematic data. Catchment areas are also represented in terms of the same raster, which allows for an efficient and flexible approach to catchment calculations.

Many of the digitising procedures described are appropriate to the available hardware and to the variable quality of primary mapping. The availability of more sophisticated technology, such as laser or raster scanning or satellite imagery, will warrant a further review of techniques of thematic data analysis in hydrology.

ACKNOWLEDGEMENTS

The author would like to thank the European Community for contributing funds towards the European Flood Study. In addition, thanks are due to the Directorate of Military Survey for advice on map projections, to David Morris who designed the segment borrowing software and to Max Beran who developed the microcomputer rastering system.

REFERENCES

Beran, M.A. 1982. Hydrology and Computer Mapping. Cartographica 19, 2. 57-61.

Gustard, A. 1983. Regional Variability of Soil Characteristics for Flood and Low Flow Estimation. Agricultural Water Management 6(1983) 255-268.

Monmonier, M.S. 1982. Computer-Assisted Cartography. Prentice-Hall, New York.

Sampson, R.J. 1975. Surface II Graphics System. Kansas Geological Survey, Kansas.

A GEOMETRIC MINE MODELING SYSTEM

Olli Karonen and Markku Tamminen
Laboratory of Information processing Science
Helsinki University of Technology, 02150 Espoo 15, Finland
Pentti Kerola, Matti Mitjonen and Esko Orivuori
Outokumpu Oy, Töölönkatu 4, 00100 Helsinki 10, Finland

ABSTRACT

We present an approach to computer aided mine design based on geometric modeling of mines and ore bodies. We discuss in some detail the computer representations and geometric processing techniques required. As a concrete example we describe the prototype system implemented for Outokumpu Oy.

INTRODUCTION

The objective of mine design is to determine good excavation alternatives taking into account physical and technological constraints and the location and quality of ore. For this we must have information on the location and form of the ore body, on the spatial distribution of its mineral contents and on the mechanical characteristics of the surrounding rock. Further we must keep track of the tunnels and excavations.

At the present mine design combines two professionally separate aspects, that of geology and that of mine engineering. The geologist explores mineral deposits and interprets vague data into assessments of ore quality. On this basis the engineer determines a detailed excavation plan together with its economic benefits and technological feasibility. Geology and engineering should be tied together with efficient communication methods and quick evaluation procedures for alternatives.

A geometric mine model serves the above purposes. It is a computer representation of a specified part of the earth's crust together with the planned and achieved excavation operations related to it. The main benefit of three-dimensional geometric models is that various evaluation procedures can be applied automatically. This is especially important as the information is continuously updated.

This paper is based on experiences gained from an experimental geometric mine modeling system implemented at the Helsinki University of Technology together with a large Finnish mining company, Outokumpu Oy. The prototype system has been applied to the small and complicated Vammala ore body in Southern Finland. The system is designed to operate on a modern workstation running the Unix operating system.

We first discuss how geometric objects related to mining can be represented in a computer and motivate our choice: the boundary representation scheme. Next we outline the prototype system describing its hardware and software environment. Finally we discuss in more depth the technically most difficult aspects of the chosen approach.

The main entity types that must be representable in a geometric mine model are the spatial distribution of rock and ore quality (natural phenomena) and the location and form of man-made objects. To describe the ore distribution we need at least the following data:

- as input the three-dimensional location of drill and bore holes and the analyzed mineral contents of corresponding samples (Figure 1)
- the type of rock in each part of space (Figure 1)
- an estimate of the three-dimensional spatial distribution of mineral contents inferred from drill holes and other measurements

The main data on man-made objects are the three-dimensional location and form of excavations, tunnels, shafts etc. (Figure 2).

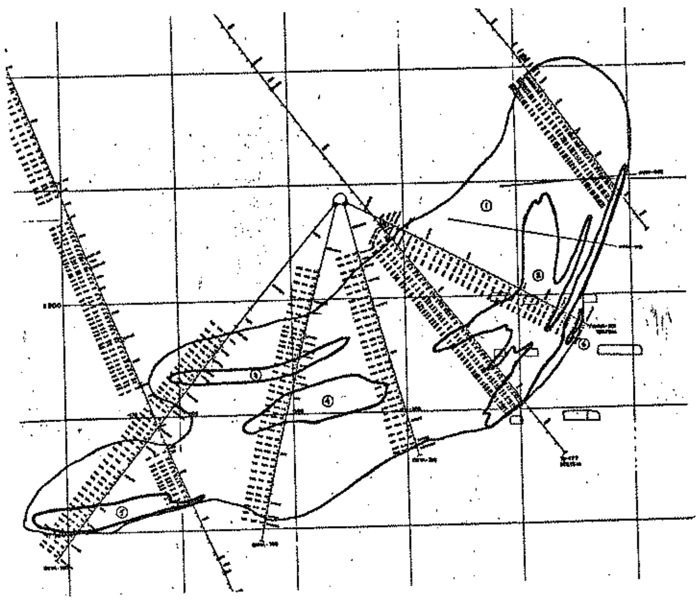


Figure 1. Section depicting drill holes and mineral borders

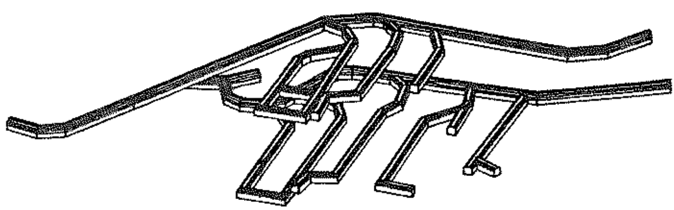


Figure 2. Axonometric view of part of Vammala tunnel network

The traditional way of representing the above data is graphical. A section map represents at correct locations drill holes, borders of rock types, borders of ore rich enough to be mined, and existing tunnels. Such a map contains both basic measured data (i.e. drill holes) and expert interpretations of a geologist (the rock type boundaries). Planar sections are easy to work on and thus they are practical as input data and as auxiliary representations also in a computerized system. At the present a small mine is represented "completely" by using several (say 10 – 30) sections.

Sectional data must be converted into volume and mass estimates. This can be done by interpreting each section as a three-dimensional slab with a thickness identical to the inter-section distance. Thus we may speak of a **slab model.** It is a viable alternative also for computer processing.

It would be easy to apply an interactive graphical system package to computerize an archive of sectional and other maps. Several such graphical systems have also been tailored for mine design. The problem is that rather few of the potential benefits of a computer can be achieved from mere graphics. It is necessary that the combined data form a **complete three-dimensional geometric model** (Requicha 1980) of the mine. To convert a purely graphical system into a complete slab model requires much programming. Also, a slab model does not well capture the geometric form of, say tunnels.

Our approach has been different: we have taken a general **solid modeling scheme** (Requicha 1980) and a corresponding core program package, **solid modeler,** which provide a complete geometric model together with various basic operations on it. Tailored interfaces make the whole a mine model.

There are several solid modeling schemes to choose from:

(1) **Spatial enumeration;** this is exemplified by the oct-tree (Meagher 1982) and similar binary voxel-trees (Tamminen 1983) that we have called **adaptive block models** because they describe space by dividing it into parallelepipeds of adaptive size (Figure 3).

(2) **Constructive solid geometry** (CSG) (Requicha 1980) describes a solid as formed by **boolean set operations** (union, intersection, difference) combining basic solids such as blocks and cylinders. Many mining operations can be modeled by set operations in a natural way; for instance, excavation is clearly a difference-operation.

(3) **Boundary models** (Mäntylä and Sulonen 1982) represent solids by their bounding surfaces further defined by more primitive constituents, vertices, edges and faces allowing any form to be described. Models can be further combined, for instance, by set operations.

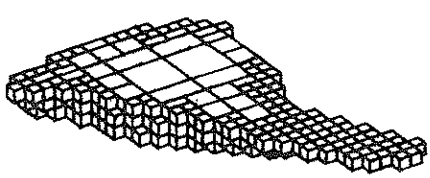


Figure 3. Adaptive block model of ore between two sections

We have chosen the boundary model as the core of the geometric mine model for the following reasons:

- tunnels, excavations etc. can be easily and precisely described
- the traditional representation of ore is by boundaries on section maps
- we can apply an operational boundary modeler, GWB (Mäntylä and Sulonen 1982) that supports the basic operations required.

We use adaptive block models as an auxiliary representation because this is a natural way to describe the spatial distribution of ore quality. Block models are compatible with the homogeneous grids used by geostatistical programs.

OVERVIEW OF THE PROTOTYPE SYSTEM

System structure and functions

At a global level the system consists of three main components:

- the geostatistical subroutine package
- the drafting subsystem, which embodies most of the traditional approach of separate section maps and utilizes methods of interacticve graphics
- the solid modeling subsystem, which makes possible three-dimensional geometric and viewing operations.

The geostatistical package uses the methods of **kriging** (Journel and Huijbregts 1978) to estimate a mineral content from input samples. For each mining area a **semivariogram** describing the spatial correlation law must first be estimated. Given a test volume (usually a small cube surrounding a given point), whose mineral contents are to be determined, a kriging program first uses spatial search to determine the, say 5-20 relevant sample points. The cross-covariance matrix is determined from the semivariogram taking into account the location of the test point and sample points. Thereafter the expected value and variance of the mineral contents of the test volume are determined similarly as in regression analysis.

When geostatistics are applied to determine the mineral contents of a two- or three-dimensional region the above operations are repeated for a grid of test points and the results are integrated over the region considered. In order to be able to apply geostatistics to a complicated boundary representation of an ore body we first convert the boundary model into an adaptive block model. We estimate each block separately and sum for a total value.

The drafting subsystem is designed to interactively process section maps (Figure 1). The basic objects known to the subsystem are drill holes and **plane regions,** both belonging to a specific x-, y-, or z-section. The plane regions (defined by their borders) may describe rock type, ore with a given quality or sections of excavations and tunnels. The data can be queried and modified on an interactive graphic terminal.

The drafting subsystem delivers to the solid modeling subsystem plane regions describing sections of rock, ore bodies and planned excavations. Also the drill hole data are used by the three-dimensional subsystem in geostatistical analyses. From the solid modeling subsystem arbitrary sections of any solids (i.e. ore bodies, excavations etc.) can be transmitted to the drafting subsystem for further interactive use.

The following are the main functions of the drafting subsystem:

- analysis–queries and graphical output of drill hole data
- the "traditional" ore estimate based on weighted averages of "drill hole sectors"
- geostatistical estimation of a point or a grid and colour scale output of the result
- specification, updating and combination of borders of plane regions
- transfer of data between sections

The solid modeling subsystem operates on three–dimensional volumes represented as boundary models. A volume is generic in that we can use it to define many different types of entities, e.g. an ore body, part of an ore body between two sections, a tunnel, the whole tunnel network or a planned excavation. Also all operations are generic; for instance we can apply sectioning or set operations or volumetric analysis to all the above entities. All volumes and other geometric entities can be viewed by the general three–dimensional graphics system Hutgraph.

The user interface of the solid modeling system must be application oriented. This means that even though the representations and basic operations are generic, the volumes are built and analyzed using concepts related to mining. For instance, a tunnel is formed by specifying a cross–section type and a center line. Similarly specific to the mining application is the automatic formation of a triangulated surface between two plane regions representing cross–sections of an ore body or a planned excavation (Figure 4). This is the interface from the two–dimensional to the three–dimensional subsystem.

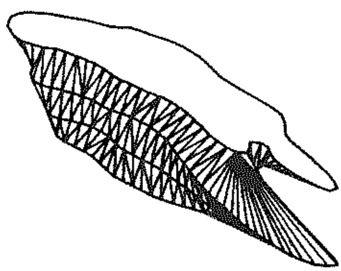


Figure 4. Formation of a surface between sections

The principal task of mine design is to correctly place a new excavation in relation to the ore body and other excavations and tunnels. Viewing the corresponding solid models gives some help. However, the main analysis of an excavation alternative consists of intersecting the model of the excavation with a similar model of the ore body and determining the amount of minerals and side material thus formed (Figure 5). For this we must be able to both perform boolean set operations and determine the three–dimensional distribution of ore quality.

The following is a list of the main functions of the solid modeling subsystem serving the above purposes:

- form a volume component by "wrapping" a surface between two plane regions and "glue" several volume components together (Figure 4)
- combine components into more complicated, bifurcated, forms
- form a volume model of a tunnel and combine tunnels into a network by using boolean union (Figure 2)

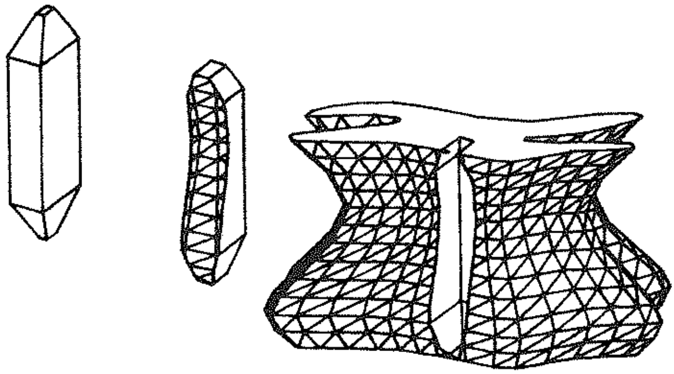


Figure 5. Modeling excavation by a boolean set operation

- form a volume model of an excavation from several cross-sections
- intersect an ore body with an excavation forming the volumes of ore and side rock (Figure 5)
- analyze geostatistically a piece of ore giving its volume, mineral contents, value etc.
- form an arbitrary section of a volume
- view an arbitrary collection of volumes and cross-sections in an appropriate projection either with or without hidden lines.

Operating environment

The geometric mine model has been designed so that it need not reside at some central computer. Using modern computer equipment it can and should reside **at the mine** on a **workstation.** Of course some parts of the system with high computational load cannot offer "real-time" response in this environment.

This far we have used a VAX 11/750 for building the prototype. During the summer it will be transported to the Motorola M68000-based SUN workstation, where most of our basic software is already operational.

The workstation approach is very flexible when combined with local networking. The prototype can be enhanced into a production system for a medium sized mine by connecting together by Ethernet a couple of independent workstations, a large enough file server and necessary plotting and printing equipment.

CORE SOLID MODELER

The layered structure of the solid modeling subsystem can be said to represent the "toolbench" approach so prevalent in an Unix environment. The basic geometric tools are provided by a small but general geometric modeler, the Geometric WorkBench (GWB). Around it we build (using the tools of Unix) a specific layer according to the needs of an application, in this case mine modeling.

GWB consists of a set of tools for creating, manipulating, storing and analyzing three-dimensional solids represented by boundary models. A model can be created by adding one-by-one vertices, edges and faces taking into account their topological (i.e. neighbourhood) relations. In GWB we can construct arbitrary valid solids by combining elementary functions called Euler operators, which automatically maintain topological integrity. The name of the operators comes from the charasteristic that each one maintains the validity of the generalized Euler formula

$$v - e + f = 2 (s - h) + r$$

between the numbers of vertices (v), egdes (e), faces (f), disconnected components (s), holes through the solid (h) and cavities in faces (r). Euler operators themselves are not user-oriented, but they can be used as primitives within higher-level operations. GWB already contains functions producing often needed solids such as rectilinear blocks and cylinders. In addition to this every application may have programs of its own to construct specialized objects. For instance forming a tunnel from its center line is a user-oriented operation of the geometric mine modeling system.

Boolean set operations are a useful but computationally difficult class of manipulations. With them a user can easily perform union, intersection or difference between solids to generate new models of great complexity.

Solid models can be named and stored on disk for later use. As the offline representation GWB uses, again, a list of Euler operators formed by "inverting" (Mäntylä and Sulonen 1982) a solid.

GWB also contains visualization and analysis functions to create graphical displays and to calculate integral properties such as the volume of a solid. Actual device independent graphical output and interaction is performed through Hutgraph, which is a "standard" three-dimensional graphics system.

A core modeler like GWB provides for generality and flexibility because it restricts in no way the form of solids and allows problem specific programming to be done at the low Euler-operation level when necessary. Despite this it contains ready made set-operations and other very high-level functions that would be extremely tedious to program. The benefit of using a general solid modeler comes from its geometric operations being applicable to any type of solid, be it an ore body, a tunnel or something else.

GEOMETRIC PROCESSING TECHNIQUES

Conversions between representations of geometry

The main conversions needed in our system are from two-dimensional sections to a three-dimensional boundary model (Figure 4) and from a boundary model to a block model. Rather much has been published on interpolating graphically between two figures (Burtnyk and Wein 1971) and even on wrapping a boundary between sections (Ganapathy and Dennehy 1982). However, often some knowledge of the relative orientation of the two sections is required.

In our case the sections should be automatically oriented with respect to each other so that their corresponding points can be connected by a triangulation. We perform this by applying the method of (Fuchs et al. 1977) to find the minimal triangulated surface between the two sections. As the method is rather time consuming we have enhanced it by determining the best orientation from the caricatures (Douglas and Peucker 1973) of the two sections. Even though we have had complicated sections to process (Figure 4) the results with the above method have, as yet, been satisfactory. In case of difficulties we let the geologist interactively specify some related point pairs in the sections.

Connecting two sections described by 80 points (giving 160 triangles, Figure 4) takes about 20 CPU seconds on our computer (VAX 11/750), which does not contain special floating point hardware. The SUN workstation requires about twice as much time for the same task.

The octree-representation, to which our block model closely corresponds, has been the object of much interest lately and entire geometric modelers have been based on it. Usually octrees are formed by either combining primitive solids (ball, cylinder , ...) by set operations or by converting plane regions described by quadtrees or pixel matrices into three-dimensional form (Yau and Srihari 1982). Rather little has been published on converting boundary models into octrees even though there exist algorithms for the corresponding two-dimensional case (Hunter and Steiglitz 1979).

In principle a conversion algorithm looks as follows:

```
function convert(var B : enclosing_box) : block_model;
begin
    if B uniformly solid then convert := form_solid_block(B)
    else if B not uniformly empty then
        convert := union(convert(first_half_of(B)),convert(second_half_of(B))
end;
```

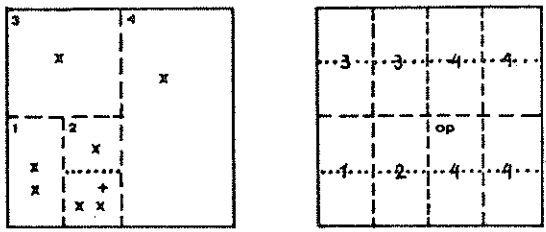
The main computational operation of this algorithm is to determine whether a parallelepiped is totally contained within the boundary model, totally outside it or intersects the boundary. Typically this operation has to be performed some tens of thousands of times with the boundary model containing hundreds of faces. Using brute force this is not practical even for large computers. To attain a resolution unit one 2,000,000th of the enclosing box we use a three-dimensional spatial index described in the next section.

Geometric directories for spatial search

There are two main uses of spatial search in the mine modeling system: In geostatistics we must find for each small test cube the 5-20 relevant drill sample points; usually relevant means closest with some extra conditions on the spatial distribution of the points. In forming the block model we must determine, whether a parallelepiped is contained in a solid. We have used the EXCELL approach to form **spatial indices** for both the above tasks.

The drill samples are stored in a permanent file organized by the point-EXCELL method (Tamminen 1982), whose essence is described by Figure 6. Using this method we can find the 10 closest points from an arbitrarily large file using only a couple of disk accesses and computing some tens of inter-point distances, on the average.

For the parallelepiped containment problem we utilize the three-dimensional EXCELL face directory described in (Mäntylä and Tamminen 1983): A parellepiped intersecting some faces is recursively subdivided until it reaches the maximal resolution. A parallelepiped intersected by no faces or of maximal resolution is classified by **ray-casting:** a semi-infinite ray is cast from its center and at the first intersection with a face we determine, whether the ray is coming from the inside or outside of the solid. The directory is, again, utilized to perform only a limited number of face/ray -intersection tests. Without it the above simple approach would not be practical. The table of Figure 7 describes the performance of block conversion on our VAX 11/750. The program has not been optimized because at the present the bottleneck of the system is still formed by geostatistical estimation.



data part   directory

x: points
o: cursor
-----: cell borders

Figure 6. Example of two-dimensional point-EXCELL concepts. The study area is divided into four data cells (=rectangles), none of which contains more than three (=bucket size) points. The directory is an array of elements, each corresponding to a rectangle of minimal size and indicating the data cell containing it. The cell corresponding to a cursor point p is retrieved by first computing the corresponding directory array index as in a regular two-dimensional matrix. The insertion of the new point (+) would require a cell division and directory doubling operation as indicated by short dashes.

resolution (m)	no blocks	CPU-time (s)	volume (1000 m3)	(error %)
8	1371	246	1383	3
16	294	84	1507	5
32	38	47	1245	13

Figure 7. Performance of block conversion for the solid of Figure 4

CONCLUSIONS

We have shown how a geometric mine modeling system can be built upon a core geometric modeler and what user interfaces and geometric processing enhancements are needed. The whole is being implemented on a modern workstation. Some parts of our system are computationally intensive and have to be further optimized before they are ready for production use on a workstation. Optimization is possible using the geometric search techniques described above.

We have chosen the boundary representation of solids. In this way we have been able to directly utilize existing general purpose software. Another relevant alternative would be the block model. However, it is unclear, whether a precise enough representation of man-made objects can be attained in this way. In all probability both representations will be required even in the production version of our system.

Geologists and mine engineers have estimated that the system will greatly increase their capability of evaluating different mining alternatives. Also, a geologist together with interactive geostatistics obtains more accurate mineral estimates than one working alone.

ACKNOWLEDGEMENTS

We thank Martti Mäntylä for his workbench and Reijo Sulonen for his comments.

REFERENCES

Burtnyk, N. and Wein, M. 1971, Computer-generated key-frame animation. J. Soc. Motion Picture and Television Engineers, 80(3)49.

Douglas, D. and Peucker, T.K. 1973, An algorithm to represent the caricature of a digitized line. Canadian Cartographer **10**(1973)4, p. 112-122.

Fuchs, H., Kedem, Z.M. and Uselton, S.P. 1977, Optimal surface construction from planar contours. CACM **20**(1977)10, p.693-702.

Ganapathy, S. and Dennehy, T.G. 1982, A new general triangulation method for planar contours. Computer Graphics **16**(1982)3, p. 69-75.

Hunter, G.M. and Steiglitz. K. 1979, Operations on Images using quad trees. IEEE Trans. PAMI **1**(1979)2, p.145-153.

Journel, A.G. and Huijbregts, Ch, J. 1978, Mining Geostatistics. Academic Press.

Meagher, D. 1982, Geometric modeling using octree encoding. Computer Graphics and Image Processing **19**, p. 129-147.

Mäntylä, M. and Sulonen, R. 1982, GWB - A Solid Modeler With Euler Operators. IEEE Computer Graphics & Applications **2**(1982)7, p. 17-31.

Mäntylä, M. and Tamminen, M. 1983, Localized set operations for solid modeling. To appear in SIGGRAPH'83 proceedings.

Requicha, A.A.G. 1980, Representations of rigid solids: theory, methods and systems. ACM Comp. Surv. **12**, p. 437-464.

Tamminen, M. 1982, The extendible cell method for closest point problems. BIT **22**(1982)1, p. 27-41.

Tamminen, M. 1983, Encoding pixel trees. Report-HTKK-TKO-B51, Helsinki University of Technology, Espoo.

Tamminen, M. and Sulonen, R. 1982, The EXCELL method for efficient geometric access to data. Proc. ACM Nineteenth Design Automation Conference, Las Vegas, p. 345-351.

Yau, M. and Srihari, S.N. 1982, A hierarchical data structure for multidimensional images. Report, Dept. of Computer Science, SUNY, Amherst, NY 14226.

COMPUTER AIDED DESIGN IN RELATION TO EARTHWORK COMPUTATIONS

Steven G. Mruk
Consultant

Synercom Technology, Inc.
10405 Corporate Drive
Sugar Land, Texas 77478

ABSTRACT

Synercom Technology, Inc., a pioneer in state-of-the-art interactive graphics systems, has developed a Route Optimization And Design System called **ROADS.** This software is general in nature and may be applied to highway design, railroad layout, and site planning. Additionally, ROADS has been used successfully for canal dredging problems and airport runway resurfacing projects. ROADS is the result of fundamental algorithms supplied by Qubit Pty., Ltd., an internationally known Australian engineering firm, and Synercom-developed interactive graphics software.

The intent of this paper is to describe ROADS' basic algorithms and file structures. A principle theme throughout this paper is to classify computational steps into three groups: (1) interactive, requiring the use of a graphics workstation; (2) interactive, executable from an alphanumeric CRT; and, (3) compute bound i.e., batch.

INTRODUCTION

Interactive graphics is the heart of the second generation of earthwork engineering computer aided design systems. Interactive graphics systems have analytical potentials far beyond those of the first generation, batch computer programs. Typical of the first generation are RDS, TERRA, AVENUE, and ROADS. These systems are briefly described as follows:

RDS

The Roadway Design System (RDS) was developed primarily under HPR Project 1-(G), 1-19-67-107 by the Division of Automation of the Texas Highway Department and the Federal Highway Administration. RDS includes more than 340 computer processes, and contains program functions beyond those that are classified as earthwork engineering. However, the graphics capability of RDS is limited to plotter output. RDS is a classic example of a batch computer program of the early '70's.

TERRA

Terrain Engineering and Retrieval for Road Alignment (TERRA) software was developed by Idan Computers, Ltd., Tel-Aviv, Israel. The TERRA system employs a batch problem oriented input language i.e., card images. Horizontal alignment concepts are based on input card formats similar to input used by the IBM Highway Design System (HIDES 1130-16.2.760, H19-0010-0). TERRA software is widely in use today and is an example of '70's code that has survived the '80's. In the author's opinion, based on benchmark results, TERRA software has the potential to make the transition from a batch oriented system to an interactive graphics system. For example, the Michigan Department of

Transportation appears to be integrating TERRA software with an interactive graphics system (Green, 1978).

AVENUE

AVENUE is a highly interactive road design program developed in Switzerland. AVENUE has evolved with characteristics that are uniquely European. In the traditional road design software found in most of the world horizontal and vertical alignments are usually defined with tangent lines. Curve and spirals are entered numerically. AVENUE is different in that centerline segments are defined randomly as either circular or straight. Given user-defined attributes, the program will compute transition curves between segments and link various segments together to form the alignment. The user is given 12 basic segments to work with for the horizontal alignment. AVENUE, if properly integrated with an interactive graphics system, will represent a powerful highway design system. AVENUE is available through WILD Heerbrugg A.G. of Switzerland.

ROADS*

ROADS or the Route Optimization And Design System has been designed to enable engineers to locate road alignments and receive rapid feedback of information. By way of plotter and printer output and by varying a small number of input parameters, the user selects the optimum alignment with minimum manual drafting and calculation requirements. An appealing aspect of ROADS is that it is highly modular, with each component program performing a unique function for the user. ROADS is both interactive and batch. The method of input varies from direct interactive input via a terminal, where only a small amount of input is required, to input via card image data files where considerable input is required.

The remainder of the paper discusses the integration of ROADS software to the Synercom Technology, Inc. INFORMAP interactive graphics system. Clearly, it is possible to integrate RDS, TERRA, AVENUE, or any other FORTRAN IV earthwork engineering system with an interactive graphics system. However, there are certain generalized interfacing techniques that if applied will result in a well-defined user oriented system. For example, user steps must be classified as interactive (requiring the use of a graphics workstation), or interactive (not requiring a graphics workstation), or simply batch. A mistake in the initial analysis and decomposition of the problem into a set of manageable program functions will result in an unwieldy system. The significance of interactivity vs. batch operations is discussed in later paragraphs.

GRAPHICS WORKSTATION

A typical interactive graphics workstation (IBM definition) consists of a graphics display screen, an alphanumeric CRT and keyboard, and a digitizing table and stylus. The components are integrated such that the operator sits in front of the digitizing table, and at eye level, can see both the alphanumerics CRT and graphics display screen. The keyboard and stylus are free to move about the table. The graphics display communicates map-like information to the user in picture form. The alphanumeric CRT and keyboard, digitizing table, and crosshair cursor (or stylus) are devices used to enter and/or retrieve terrain information interactively.

*ROADS fundamental algorithms were developed primarily by Roger Woolley of Qubit Pty., Ltd., 18 Prawse Street, West Perth, West Australia 6005. ROADS is a modified version of the Qubit Pty. Ltd. suite of ROADS programs.

Synercom "INFORMAP" software operates as a monitor and controls queries and data filing. Auxiliary to the graphics workstation and INFORMAP are MAP/IN, INFORM, COGO/I, CAP/IN, and PAC/IN. These are subsystems of INFORMAP which perform as the human interface for production and management of geographically organized information. These subsystems act as "pre" and "post" processors for route optimization and design software. Additionally, INFORMAP supports multiple schema, menus, rules, macros, etc. A full discussion of INFORMAP is beyond the scope of this paper.

MAP/IN

MAP/IN is an automated mapping and facilities data acquisition system that provides a means of mapping control and standardization, while acquiring and organizing a global data base of detailed information. MAP/IN supports production of finished maps rapidly and efficiently while reducing mapping costs. With MAP/IN, standard, high-quality maps are produced in a fraction of the time it would take trained draftsmen using conventional methods. Changes and updates, which affect a number of related maps, are effected with cost savings. Interactive graphics provides the ability to see the results of a modification or addition to a map as the changes are made and adds to the user's confidence and productivity.

INFORM

INFORM is designed to operate in an integrated manner with MAP/IN. This combination provides the user with a method of acquiring and maintaining a geographically-oriented data base. INFORM supports management and mapping functions that utilize the data base. This reporting capability is one of the principal benefits of the system and justification for data base creation. INFORM is a real-time management information system. It utilizes a geographically organized data base to define the area being managed. The data base may be selectively accessed to provide concise, timely, and up-to-date information in graphic or attribute form. For example, a map or display of selected geographic and facilities data contained within a specified area, or a cross-tabulation report on selected facilities.

COGO/I

The Coordinate Geometry Interactive System (COGO/I) is a subsystem of MAP/IN and INFORM. COGO/I supports the detailed, special requirements of the civil engineer and surveyor. The capabilities incorporated in COGO/I support entry of facility data. Typical applications are the location of facilities at a specified distance from a property line or from another facility. COGO/I may also be used to solve simple engineering problems such as the location of an element from a known monument, at a given bearing and distance.

CAP/IN

Developed jointly by Synercom and WILD Heerbrugg A.G. of Switzerland to meet the specialized requirements of photogrammetric and cartographic operations, CAP/IN combines full stereo mapping with INFORMAP capabilities of interactive mapping and geographic data base management. The system provides on-line connection of photogrammetric stereoplotters. Special consideration is given to the fact that compiling a stereo model is different from digitizing maps on a digitizer.

PAC/IN

PAC/IN is interactive PLAN, PROFILE, and CROSS-SECTION software. PAC/IN software is a comprehensive set of INFORMAP macros and FORTRAN programs. The user interacts with PAC/IN through: (1) digital terrain data record image format descriptions; (2) schema and grid parameter definitions; (3) drawing sheet layout; (4) data entry subsystem; (5) cross-section generation; (6) profile generation; (7) editing subsystem; and, (8) data extraction.

ROADS Revisited

Computational earthwork programs in an interactive graphics environment may be divided into four functional groups: (1) ground model development; (2) alignment design; (3) volumetric calculations; and, (4) interface programs.

Ground model development programs consist of partitioning the data into manageable geographic units, sorting the data within each unit, and interpolating z values for arbitrary planimetric points. Alignment design programs can be classified into two types: horizontal and vertical. The vertical alignment group includes template definition. Volumetric programs have two entry points. First, it is desirable to calculate volumes for templates on an alignment and second, it is essential that volumes be computable from two arbitrary surfaces for any closed polygon. Finally, interface programs transmit data to/from the INFORMAP interactive graphics system for rapid display and editing i.e., PAC/IN subsystem.

The standard procedure to be followed by engineers in designing a road system is dictated by the nature of problem. Aside from building the ground model, interpolation is the most time-consuming step. Thus, it is important the interpolation be cleverly coded so that run times are minimized thereby assuring reasonably fast response times for user interactivity. Additionally, the system must be efficient in terms of disk space. To accommodate both problems, data organization for interpolation and disk space, ROADS software portions the geographic area of interest into bands. The data within each band of the digital ground model (DGM) is integerized and sorted. The user has complete control of the bandwidth according to the following formulae:

a. Minimum Number of Records

$$\text{NREC (min)} = \frac{\text{N(max)-N(min)}}{\text{Bandwidth}} + 2$$

b. Maximum Number of Records

$$\text{NREC (max)} = 5\,(\text{NREC (min)} - 1)) + 1$$

where: N (max) = Maximum Northing Coordinate of DGM
N (min) = Minimum Northing Coordinate of DGM

The maximum size of the DGM is limited by integer limit of 32767, but selection of units usually overcomes this limitation.

The record for any location can be determined as follows:

$$\text{NREC} = \frac{\text{N-N (min)}}{\text{Bandwidth}} + 1$$

where: N = Northing Coordinate of Point

NOTE: The above formula assume that the north direction is the primary or longest direction of the model. If this is not the case, N then refers to the easting coordinate. Additionally, the program clips the designated area by one bandwidth in the easting direction if the DEM file is created E-W or in the northing direction if created N-S. This enables record 1 to be used for storing origin dato for subsequent accesses of the file. The DGM is limited to 100 ground level points per record plus four overflow records therefore giving a limit per bandwidth strip of 500 points.

Bandwidth File (BW...)

All doto in the DGM is stored in integer form. The file record length is 606 bytes, that is 303 16 bit integer words. Data is stored in the following locations:

a. First Record:

IRRAY (1,4)	=	easting min.	
IRRAY (5,8)	=	northing min.	
IRRAY (9,12)	=	easting max.	
IRRAY (13,16)	=	northing max.	Store as Double Precision
IRRAY (17,20)	=	vertical datum	Values
IRRAY (21,24)	=	bandwidth	
IRRAY (25,28)	=	total points on file	
IRRAY (301)	=	file length	
IRRAY (302) 0	=	E-W orientation	
1	=	N-S orientation	
IRRAY (303)	=	highest record assessed	

b. Subsequent Records:

IRRAY (1), (4), (7), to (298)	=	east increment
IRRAY (2), (5), (8), to (299)	=	north increment
IRRAY (3), (6), (9), to (300)	=	elevation
IRRAY (301)	=	overflow record location
IRRAY (302)	=	number of elements of data in the record, being 3 elements of data per ground level point.

Interpolation

ROADS supports two methods of interpolotion, a least squares method and a triangulation method. Since each method of interpolation is confined to a subroutine, the user has the option of selecting a particular method. Additionally, the user may substitute this subroutine with his own algorithm of interpolation.

In terms of program efficiency, recall that the natural elevations were partitioned into bands and sorted from top to bottom within each band. This facilitates efficient and fast run-time interpolations. To clarify this consider the opposite. In the case without partitioned and sorted bands, the program would have to read the entire DGM file for each poss in collecting the neorest points for interpolation. This is a laborous step in terms of CPU processing and disk I/O. By partitioning and sarting, the program need only check a limited number of bands depending on the width of the search area in question.

Horizontal Alignment

This algorithm calculates road centerline coordinates at station intervals bosed on the horizontal alignment data provided. This includes the calculation of circular curves, spiral transition curves and super elevation. The algorithm can be used to calculate a single alignment, with or without circular curves and with or without spiral transitions ta curves. Where spiral transitions are required, the length of transition must be specified numerically and this length will be applied at both ends of the circular curve. Super elevation will be calculated if a design speed is specified. The super elevation is, however, not a smooth transition from a standard 3% crossfall but is determined from the horizontal level. A manual smoothing adjustment at an interactive graphics workstation is therefore necessary within the transition length. This will result in a very slight error in the earthwork's quantities.

Cross-Section File (XS...)

This is a 50 single precision word (100 16-bit integer words) file into which is written the basic cross-section data.

IRRAY (1-2)	=	station
IRRAY (3-4)	=	easting
IRRAY (5-6)	=	northing
IRRAY (7-8)	=	azimuth (radians)
IRRAY (9-100)	=	offset and elevation pairs

Elements IRRAY (9-100) contain offsets and elevation as determined by interpolation, with the centerline offset (value = 0) beginning in IRRAY (54). Thus, the cross-section file will support 11 points to the left and right of the centerline.

SETOUT File (SO...)

This is a 5 double precision word (20 16-bit integers) file used when super elevation of roads is to be applied. This file stores crossfall information. It also stores the basic cross-section data in double precision format which may be required for setting out purposes. The following data is stored in each record:

IRRAY (1-4)	=	station
IRRAY (5-8)	=	easting
IRRAY (9-12)	=	northing
IRRAY (13-16)	=	azimuth (radians)
IRRAY (17-20)	=	super elevation % slope

NOTE: This data file may be omitted if super elevation is not a design consideration.

Formula Used:

a. Circular Curves:

i) arc length
$ARC = \Theta c R$

ii) tangent length
$T = R.\tan \frac{\Theta c}{2}$

iii) external ordinate distance
$E = R\,(\sec \frac{\Theta c}{2} - 1)$

b. Spiral Transition Curves

i) spiral deflection angle

$$\Theta s = \frac{Ls}{2.R}$$

ii) descissa distance

$$x = Ls - \frac{Ls^3}{40.R^2} - \frac{Ls^5}{3456.R^4} - \frac{Ls^7}{599040.R^6}$$

iii) ordinate distance

$$y = \frac{Ls^2}{6.R} - \frac{Ls^4}{336.R^3} + \frac{Ls^6}{42240.R^5} - \frac{Ls^8}{9676800.R^7}$$

iv) long chord length of spiral

$$CHORD = x^2 + y^2$$

v) deflection angle of SC from main tangent

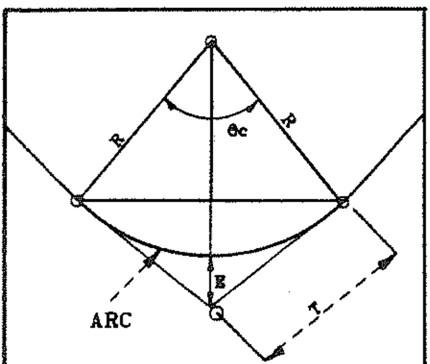
$$\delta_s = \frac{\Theta s}{3} + \frac{\Theta s^3}{105} + \frac{\Theta s^5}{5997} - \frac{\Theta s^7}{198700}$$

vi) tangential shift

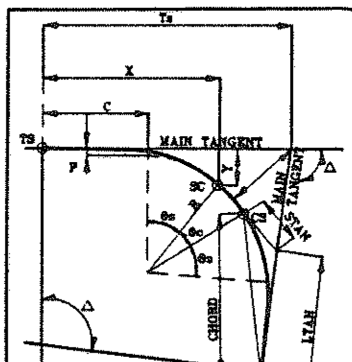
$$c = \frac{Ls}{2} - \frac{Ls^3}{240.R^2} + \frac{Ls^5}{34560.R^4} - \frac{Ls^7}{8386560.R^6} + \frac{Ls^9}{31585075.R^8}$$

vii) rodial shift

$$s = \frac{Ls^2}{24.R} - \frac{Ls^4}{2688.R^3} + \frac{Ls^6}{506880.R^5} - \frac{Ls^8}{154828800.R^7}$$



CIRCULAR CURVE



TRANSITION CURVE

viii) short tangent of spiral

$$STAN = \sqrt{\left(\frac{Y}{\tan \Theta s}\right)^2 + y^2}$$

ix) long tangent of spiral

$$LTAN = x - \frac{Y}{\tan \Theta s}$$

x) main tangent length

$$Ts = (R + s) \cdot \tan \frac{\Delta}{2} + c$$

xi) main external ordinate distance

$$Es = (R + s) \cdot \sec \frac{\Delta}{2} + s$$

where: L_s = Length of spiral transition
R = Radius of circular curve
Δ = Intersection angle of the main tangents

Vertical Alignment

This algorithm calculates the vertical alignment of a road centerline at station intervals based on the cross-section increments. Calculation is commenced at a specified elevation and is based on station and elevation values being specified at each vertical curve location and at the end of the alignment. The interactive graphics workstation may be used to supply station and elevation values. The ability of PAC/IN software to edit from a profile view is essential to the design of the vertical alignment. The length of the verticol curve may be either specified or can be computed by the program based on specified speed, friction, and sight-distance data.

Vertical Alignment File (VA...)

The resulting vertical alignment data is stored in a single data file. This file consists of vertical curve data for further use with template design considerations. This file is a 3 word single precision (6 integer words) file constructed as follows:

IRRAY (1-2)	=	station
IRRAY (3-4)	=	elevation
IRRAY (5-6)	=	design cross-section template number (determined after templates are related to the alignment in a computational step)

Formula Used

Sag curve for headlight sight distance

$$L = \frac{A.D.^2}{150+3.5D}$$

where: L = Length of vertical curve
A = Algebraic difference in grades (%)
D = Heodlight sight distonce

NOTE: L is rounded up to the next specified round-off value
L is greater than the sight distance required

Summit curve for stopping sight distance

$$D = 2\left(0.7V + \frac{V^2}{254f}\right)$$

$$L = \frac{D^2 A}{200(\sqrt{h_1}+\sqrt{h_2})^2}$$

where:			
	L	=	Length of vertical curve
	A	=	Algebraic difference in grades (%)
	D	=	Stopping sight distance
	V	=	Design speed of vehicle in km/hour
	f	=	Skidding coefficient of friction (normally 0.35)
	h_1	=	Height of eye above road (normally 1.15m)
	h_2	=	Height of object above road (normally 1.15m if another vehicle or 0.2m if any object on ground)

NOTE: L is rounded up to the next specified round-off value
L is greater than the sight distance required

Rate of change of grade

$$K = \frac{L}{A}$$

External distance

$$E = \frac{LA}{800}$$

Templates

This algorithm sets up a menu of typical design cross-section templates which will apply along the alignment to be analyzed. A template is required for every variation required in cross-sectional shape along the alignment. The template defined may include berms on the embankments based on maximum embankment heights provided. Curbs and open channel drains must be defined. Additionally, right and left embankment slopes must also be defined. The elevation datum used for each cross-section is the edge of road, defined as the first offset point either side of the centerline. This edge of road elevation, when determined, is based on the road centerline and the crossfall applicable.

Template File (TS...)

In ROADS a template will support up to eight offset points on either side of the centerline. The data is stored in 42 single precision word records as follows:

i) offsets and level differences left of centerline

offset	(1)	=	word (16);	diff	(1)	=	word (15)
	(2)	=	(14);		(2)	=	(13)
	(3)	=	(12);		(3)	=	(11)
	(4)	=	(10);		(4)	=	(9)
	(5)	=	(8);		(5)	=	(7)
	(6)	=	(6);		(6)	=	(5)
	(7)	=	(4);		(7)	=	(3)
	(8)	=	(2);		(8)	=	(1)

ii) offsets and level differences right of centerline

offset	(1)	=	word (19);	diff	(1)	=	word (20)

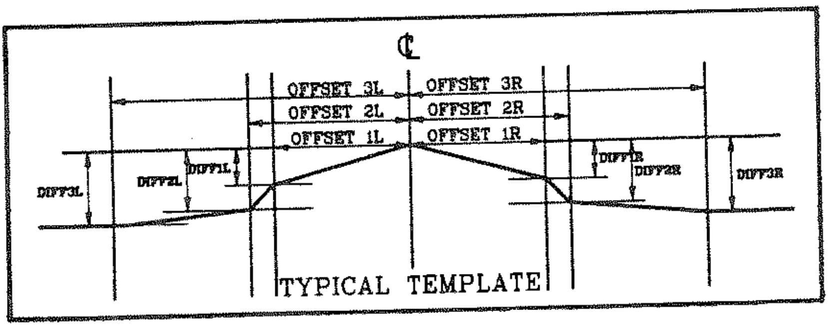
(2)	=	(21);	(2)	=	(22)
(3)	=	(23);	(3)	=	(24)
(4)	=	(25);	(4)	=	(26)
(5)	=	(27);	(5)	=	(28)
(6)	=	(29);	(6)	=	(30)
(7)	=	(31);	(7)	=	(32)
(8)	=	(33);	(8)	=	(34)

iii) no. points left = word (35)
no. points right = right (36)

iv) left cut slope = word (37)
left fill slope = word (38)
berm width = word (39)
height to berm = word (40)
right cut slope = word (41)
right fill slope = (42)

v) word (17) and word (18) = 0.0

NOTE: Diff is expressed as a function of the crossfall from the centerline with the centerline itself being regarded as zero.



Computational Step

This algorithm forms the final design cross-section based on existing ground cross-sections and typical cross-section templates. The user can specify individual templates for the four cases: cut/cut, fill/fill, cut/fill, and fill/cut. The algorithm bases its original cross-section template selection on a specified estimated width from the road centerline to the edge of the earthwork's formation. It then reiterates the process based with a more refined calculated distance and shape to select the final template. There are several additional variations that can be applied to the finally formed up cross-section shape. The variations are as follows:

a. Variable Embankment Slope - May be specified if slope is to be varied depending on height. There is allowance for two slopes to be inserted and the height to the change in slope is taken as the height used to specify the berm location for the template.

b. Ditch Carry Through - The maximum height of fill through which an open channel drain or ditch is to be carried through between two areas of cut may be specified.

c. Minimum and Maximum Verge Slope - The minimum and maximum slope of road verge may be specified. This will limit the required formation to these limits within the verge width specified. Outside the verge limits, the normal slopes as per the template are adopted.

d. Fill Height to Width Ratio - A factor may be specified indicating a ratio of embankment fill width to height. The width will be increased if necessary to suit this factor.

e. Superelevated Curve Widening - A distance may be specified which will be added to the offset to the edge of road line at the inside of horizontal curves. The "bulge" starts at the tangent to spiral and stops at the spiral to tangent points.

f. Widening Option - This option allows the user to select up to 10 elements on the correct template to be varied linearly between the NFROM and NTO sections.

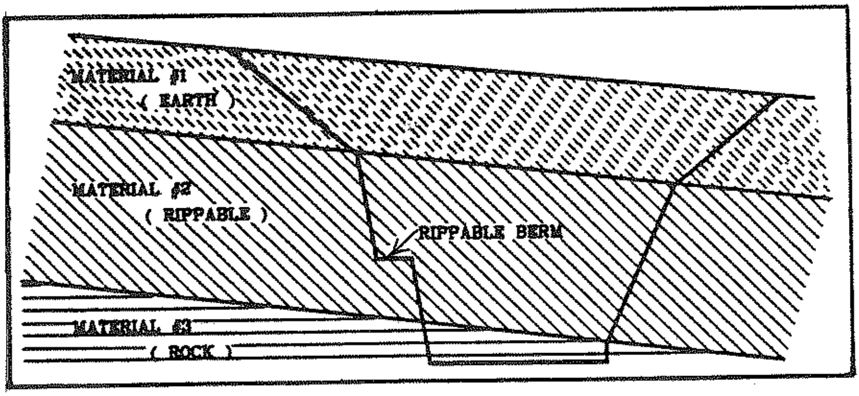
Design Section File (DS...)

The data for the finally formed cross-sections are stored in a 125 single precision word file and is stored as follows:

word	(1)	=	offset distance from centerline to junction of natural ground and design formation at left side
word	(2)	=	elevation of above point
word	(3)	=	offset distance to natural ground level between above point and the centerline
word	(4)	=	elevation of above natural ground point
word	(3)	to	
	(46)	=	are used to store natural ground offset and elevations as per (2) and (3) which occur within the formation, with the two words after the last offset representing the right side junction as per (1) and (2)
word	(47)	=	template number
word	(48)	=	0.0
word	(49)	=	azimuth of centerline point (radius)
word	(50)	=	number of final section data elements (each point consisting of an offset and elevation element)
word	(51)	=	as per (1)
word	(52)	=	as per (2) unless design does not meet existing ground
word	(53)	to	
	(120)	=	are used to store final section offsets and elevations as for (3) to (46)
word	(121)	=	centerline easting
word	(122)	=	centerline northing
word	(123)	=	0.0
word	(124)	=	0.0
word	(125)	=	the word number of the last final cross-section data element (between words 53 and 120), plus 1

Geological Strata

This step allows for data entry to determine the depth of the rippable and/or rock strata to enable different engineering parameters to be used as the alignment passes through areas of cut. It should be noted that the program makes the assumption that the depth is constant over the whole section. The program accommodates rippable and rock berms and side slopes.



Geological Strato File (VO...)

The file produced by this program is a collection of 15 single precision word records as follows:

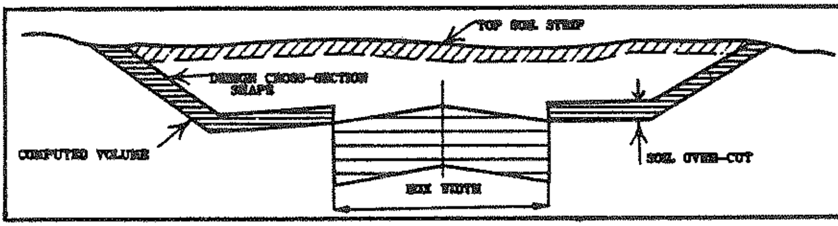
word	(1)	=	station
word	(2)	=	rippable depth
word	(3)	=	rock depth
word	(4)	=	rippable slope
word	(5)	=	rock slope
word	(6)	=	rippable berm
word	(7)	=	rock berm
word	(8)	=	minimum rippable depth
word	(9)	=	minimum rock depth
word	(10)- (15)	=	reserved

Volumetrics

This algorithm calculates the material volumes from data obtained from the design cross-section file (DS...) taking into account the following:

a. bulking factor in percentage form for the cut/fill material
b. road box width
c. road base thickness
d. top soil to be stripped prior to excavation, stockpiled and not used as fill
e. over-cut for replacement with top soil
f. the various side slopes through the different types of materials as stored on the VO... file.

These factors can be varied between various sections of the alignment. The resulting volumes are based on both cut and fill areas in each cross-section. The fill volumes are adjusted by the program with the bulking factor specified. The over-cut area is at this stage calculated on a projected flat surface area and not on the actual area of the sloping surface. This results in o minor error in the volume of cut calculated (i.e., actual greater than calculated). Output from this algorithm is a volumetric report with summaries for cut and fill at each cross-section in addition to cumulative totals. The algorithm uses the average end area method of computation.



Site Planning

This algorithm computes cross-sections aver a given polygonal area. The area defined by the vertex coordinates are entered as (x,y) pairs. This data can be entered graphically through the stylus and digitizing table at a graphics warkstatian, or manually from the keyboard. Cross-sections are generated at given intervals and the distance between points on each sectian is alsa user definable. Additianal crass-sectians are autamatically generated at polygon vertices in relation to an imaginary centerline. The cross-sectioned area is then applied ta two DGMs for interpolation. The average end area formula is applied ta the crass-sections, permitting volumetric calculations. Additionally, this algorithm supparts a linear interpalation for areas defined by (x,y,z) triples. It is thus possible to generate a reasonable surface for a dump or stockpile base when no data is availoble far the underside of the pile.

Site Planning File (SP...)

The computed data points are written to a 20 16-bit ward (5 double precision) file with the first double precision word being equivalenced to 4 integer words used for identification records and its remaining 4 allocated to x, y, z_1, and z_2 values:

DP word (1)	=	identification
DP word (2)	=	x coordinate
DP word (3)	=	y coordinate
DP word (4)	=	z coordinote (first model)
DP word (5)	=	z coordinate (second model)

Interactivity

The cohesive force that ties this system together and distinguishes it from its predecessor generation, batch computer programs, is the graphics workstation. Virtually any input for ROADS thot can be represented by a point can be entered through the graphics workstotion i.e., digitized input. For example, rather than define an alignment, run a program, etc., it is possible to graphically digitize several alignments and loop through algorithmic steps, thereby choosing an optimal alignment, interactively, in a single pass. Programs not requiring a graphics workstation may be run from an alphanumeric CRT i.e., minor edits to input files. The "trick" to the entire system is to partition and sort model data, in a batch mode, prior to the highly interactive design steps.

Variables NFROM and NTO used throughout ROADS are sequenced integers that are mapped onto the real valued station space. Variable IGROUP is an indicator that relates o porticulor subset of station space together. For example, NFROM = 1, NTO = 50, IGROUP = 1 relates the first 50 cross-sections to group 1 i.e., natural elevation. The set, NFROM = 1, NTO =50, IGROUP = 2 relates the same 50 cross-sections to o second group i.e., designed templates. With these three variables (NFROM, NTO, and IGROUP) the user has complete interactive control over any area of design at any given design step. This data is passed between INFORMAP and ROADS in 80 byte ASCII (AS...) card image files (Mruk, 1982).

SUMMARY

The components of a route optimization and design system are as follows: GWS-Graphics Workstation; MAP/IN - generalized data base mapping software; INFORM - reporting capabilities; COGO/I - interactive coordinate geometry system; CAP/IN - cartographic and photogrammetric software; PAC/IN - plan profile and cross-section subsystem; model definition; interpoloting; horizontal and verticol olignment; template definition; computational step-relate template to alignment; geological strata definitions; and volumetrics. In addition to route optimization and design the system will cross-section an arbitrary polygon for site planning.

File Summary

BW...	Bandwidth File	DS...	Design Section File
XS...	Cross-section File	VO...	Geological Strata
SO...	Setout File	SP...	Site Planning File
VA...	Vertical Alignment File	AS...	ASCII Interchange File
TS...	Template File		

CONCLUSION

The second generation of earthwork engineering systems involves interactive computer graphics. In the analysis of earthwork computations, there are three classes of operations: (1) interoctive, requiring the use of a graphics workstation i.e., alignment definitions, polygonal boundaries, etc.; (2) interactive, not requiring the use of o graphics workstation i.e., algorithmic computations, minor edits to alignment data, etc.; and, (3) compute bound i.e., model definition ond sorting. Naturally, the three classifications structure the problem so that it is easily implemented in an interactive environment. It is reosonable to assume that without the data being partitioned into bands and sorted in the early stages of model development, the entire process would become compute bound.

ACKNOWLEDGEMENTS

The author is indebted to Mr. Roger Woolley of Qubit Pty. Ltd. for the material on ROADS presented in this paper. Mr. Woolley traveled half-way around the world to come to Houston and install ROADS software. In three short weeks Mr. Woolley had his Prime version converted and aperotional an a PDP 11/70. Naturally any omissions, altered files, or outright errors in the interpretation of Mr. Woolley's original suite of ROADS programs are the responsibility af the author.

REFERENCES

1. Green, R.R., "The Role of the Numeric Ground Image System Within The Michigan Automated Transportation Engineering System", ACSM-ASP, Digital Terrain Models Sympasium, May, 1978.

2. Mruk, S.G., "System Analysis and Design of An Interactive Digital Terrain Modeling System", ACSM-ASP, Fall Technical Paper, September, 1982.

AUTOMATED DATABASE CAPTURE FOR CADD

Richard N. Stover
CADD/CAM Advisors, Inc.
5900 Princess Garden Parkway, Suite 601
Lantham, Maryland 20706 U.S.A.

ABSTRACT

One not so obvious problem with CADD systems is building a data base. CADD vendors rarely mention this because, depending on the application, users can spend years to capture a database. During this time period, productive work for which the system was purchased is delayed. This author and others have discussed optical scanning methods for database capture and while some potential solutions exist, the real problems have not been defined. The intent of this paper is to properly define problems which can aid users in selecting systems configurations for their needs.

INTRODUCTION

All CADD and CAD/CAM systems require databases of current and archival information from which the user can produce new designs, new drawings and support CAM operations, such as NC programming.

A Database for operation of a CADD or CAD/CAM system can be as simple as standard notes and text for placing on mechanical drawings to a complete set of maps for a municipality. Government agencies such as Defense Mapping probably have the most extreme database requirements, those necessary for global mapping. Between these extremes lie those applications which require libraries of symbols such as electronic (IC's, PC's), mechanical (gears, valves, screws, etc.) and architectural (fixtures, furnishings, standard room layouts, etc.). In addition, archival information can be used as a base for new designs and drawings. In manufacturing operations, Group Technology (that function which provides grouping of designs, parts, components, subassemblies and assemblies which are similar in design and manufacturing methods) needs the database of all previously designed and manufactured parts.

This author has discussed optical scanning methods of capturing databases for CADD and CAD/CAM systems. (1) These methods include line following and raster scanning with raster-to-vector conversion. The cost effectiveness of these methods for most CADD and CAD/CAM applications depends heavily on the amount of intelligence which can be automatically attached to the captured data.

CADD and CAD/CAM databases should contain intelligent data which will allow retrieval of useable pieces of information. This intelligence should fit the requirements of database management.

DATABASE MANAGEMENT

Database management requires two fundamental properties of the data: associativity and connectivity.

Associativity is the ability to link non-graphic text or attribute data (descriptive information) with specific graphic items. Such items could be components which make up a larger assembly. The non-graphic information can be visible and shown on a drawing or invisible and stored in a separate file. If the attribute data is stored separately, pointers must be provided back to the specific graphic items.

Associativity allows one to select a symbol and receive all related attribute information. Thus, it is possible to store all cost and ordering information with a specific component if so desired.

Connectivity is the ability to distinguish which graphic items are linked to each other simply by selecting a specific item. For example, on a utility map it is important to know which line connects to distribution transformer at one end and to a building entrance at the other end.

These intelligent properties also enable a CAD/CAM system to update corresponding text and graphics files to reflect revisions made to either.

Some automated digitizing vendors are developing post processors for character and symbol recognition which will provide some of the needed intelligence. The success of such tools will determine the cost effectiveness for most applications. (In the series referenced previously, this author inferred that applications requiring digitizing of long, non-straight lines such as found on contour maps could be cost effective for many automated systems while applications requiring digitizing of short, straight lines such as found on simple mechanical drawings would not be very cost effective. Single vector digitizing on an automated system may not be cost effective, while multiple vector digitizing can be. Contour lines represent the latter.)

CHARACTER RECOGNITION

The lack of successful post processors for character recognition has almost become legendary. Aside from the problems of non-uniform lettering on many drawings and maps, it is necessary to handle text in any orientation. Lettering which touches lines tends to fowl most attempts at recognition. Other similar kinds of problems have caused the task to be difficult even when Leroy™ lettering and Varityped™ lettering has been used. Nevertheless solutions are now being promised and some elementary demonstrations are being made. Some of the better approaches use a "teach mode" where an operator intervenes and "teaches" the recognition system during the first pass through the alphabet. The recognition system then stores the image (usually connectivity information with some combination of strokes or vectors, nodes or intersections, termini, directions and lengths of vectors). The systems which provide the best vectorized data will probably be the most successful at recognition since all characteristics must be known to properly parse an image of any kind.

A less difficult task should be that of symbol recognition and some degree of inferencing (deducing certain parameters or characteristics based on other known parameters). A seemingly simple example of this is to determine the end points of a long straight line which is usually output as short straight line segments from the raster to vector processing.

Other capabilities include line width recognition, testing intersection angles and adjusting to preset angles all those within certain tolerances. This allows right angles to be tagged and adjusted as right angles so that rectangles can be determined and produced. Similarly equilateral triangles can be determined and produced.

Additional capabilities include testing of arcs and circles and determining center points and radii.

Such problems have not been solved by many of the vendors. Only recently have some been addressing these kinds of problems. Solutions are being promised although few are currently demonstrating results.

Other capabilities to be addressed by some vendors include interfacing of the vectorized formats to some CAD/CAM systems. Without the interfaces to the CAD/CAM systems data bases, the purpose of the data capture systems will not be well served.

DOCUMENT QUALITY

This brings up one other point concerning the utilization of optical scanning data capture systems. The quality of the document to be scanned is all important to the data capture process. At the present time only high quality documents which have been inked and/or photographically reproduced can reasonably be expected to be scanned and captured. Pencil drawings, blueprints and even some microforms are of questionable quality. Even though such documents can be vectorized, the amount of interactive editing needed to render them useful can add greatly to the cost of capture.

One vendor has partially overcome this by preprocessing and thinning the raster data but this requires another step which may or may not be desirable. Further developments with the scanning hardware and the treatment of line edge transitions should enhance the ability to use less perfect documents in the future.

AUTOMATED MAPPING ECONOMICS

As mentioned earlier in this paper, CADD systems for mapping or automated mapping systems have the most extreme requirements in building a database.

Typical of these requirements is one large city which began to build its database in 1976. Recently, after five or more years of effort, the bulk of this digitizing project was completed. During this peak five year period, 20 digitizing CAD workstations operating 3 shifts per day were employed to perform this task. That translates into 2,000 hrs/year/shift/workstation x 3 shifts x 20 workstations x 5 years = 500,000 hrs. Assuming a conservative loaded hourly cost of $15 per hour (direct labor plus fringes and overhead of supervisory system support and clerical support), this totals $7,500,00. excluding system cost.

The system costs are estimated to have been $2,000,000. Assuming a 5 year payback, the depreciation costs are $400,000. per year. Assuming interest costs at 12% per annum and annual maintenance costs at 12%, the total yearly costs for equipment were $400,000. + $240,000. + $240,000. = $880,000./year x 5 years = $4,400,000.

Total operating and equipment costs for 5 years only = $11,900,000.

Total yearly cost $2,380,000. Few cities or municipalities can afford such costs of preparing a database.

Even if the costs of an automated digitizing system were the same ($2 million), but assuming the system could be operated by 5 persons per shift for scanning and editing the documents, the savings would be 2000 hours/person/shift x 15 persons x 3 shifts x $15./hour = $1,350,000./year x 5 years = $6,750,000.

The total project cost is now reduced to $5,150,000. At approximated $1,000,000. per year, the project may be more feasible for other cities. Also with decreasing systems costs, even less should be possible.

While not all digitizing can be automated, the technology may now be capable of handling larger shares of most requirements. Careful analysis of requirements is called for.

SUMMARY

So what are the real problems of automated digitizing systems. Two main types are usually mentioned.

1. Insufficient data base intelligence for many applications.

2. Inability to scan and utilize commonly available maps.

Insufficient data base intelligence can mean different things for different applications. Scanning and digitizing a map of contour lines with numeric elevation values inserted at breaks in the contour lines can be handled in several ways. One method is to scan and capture lines and numbers as graphic information. Using the edit capabilities of some CAD/CAM systems, the numbers can be deleted from the graphics file and inserted via keyboard into an attribute file with appropriate pointers. At the same time,the line gap left where the graphic number was removed can be repaired and replaced with a solid line. We now have a manageable data base of contour information. The elevation information in the attribute file is <u>associated</u> with the correct contour line and each contour line is correctly <u>linked</u>. Of course, it would be desirable to use character recognition to automatically recognize the elevations and enter the numbers into the attribute files. After that operation, the line gap must still be repaired either automatically or by manual interaction. A number of interactive techniques are also available for automatic tagging of contour elevations but these do not involve character recognition.

A different application may involve a 3-D piping diagram which may include piping, valves and piping support members. Unless the valves are coded by some recognizable symbols, piping, valves and supports will all be recorded as line data. Assuming the valves could be recognized in the correct order of connectivity, the support members would still not be distinguished from the piping. Enter the CADD/CAM editing capabilities to provide the identification to each support member (line at a time). But is this needed amount of interaction cost effective? Chances are that is it not. Such problems increase as the complexity of information on the documents increases.

The second problem relates to the quality of documents normally used by humans. Blueprints, coffee-stained and wrinkled drawings are not good candidates for scanning. Even Xerox copies are usually not adequate.

If you examine any linework of blueprints, Xerox copies or even original pencil copies under magnification, you will see a granular structure across the entire width of the line. The human eye in conjunction with an intelligent brain tends to integrate this information into a thick normal looking line with rather finite edges. A scanner with 0.001 inch or even 0.005 inch resolution will resolve all of these dots. Raster to vector conversion has not reached the sophistication of a human brain to integrate all of this data into a line of finite width. Therefore, the amount of processing time becomes unreasonable and the resulting product may look unusable.

As mentioned, most scanning has employed either ink on film or photographically reproduced copies to provide good input data. This is not always practical. Even so, line breaks which are not evident to the naked eye can cause problems. The only solutions are to employ more electro-optical procedures to integrate and enhance the scanned document image and additional software techniques to reduce the dependence of the vectorizing algorithms on the line quality.

The following is a capsule summary of each vendor's known progress since we published the last series of articles. (1). Note that we have only included known vendors who offer complete vector system solutions for CADD. This does not include image raster scanners and raster image processing systems.

VENDORS

ANA Tech

This company being one of the more recent entries has taken advantage of some of the known inadequacies and may possess one of the better systems. As of this writing, ANA Tech has completed a prototype system with high speed vectorization capabilities.

But the vectors created by the system (a trapezoidal format) may not conform to the basic requirements of most CAD/CAM systems. (Most CAD/CAM systems use a basic generic pen plotting format.) Benchmarks are now being performed and we should soon have a better evaluation of ANA Tech's ability to provide a structured intelligent format for input to a CAD/CAM system.

With regard to system features and capabilities, ANA Tech has built a bit slice parallel processing hardware system which takes the output from a scanner (ANA Tech does not manufacture scanners) and produces the trapezoidal vector output. The hardware processor accepts raster output from the scanner in real time at speeds up to 10 Megahertz. Currently using an EOCOM scanner an "E" size drawing can be scanned and processed to the trapezoidal format in one minute.

Note that this is not the total processing time for a drawing. A VAX 11/750 is used to post process the trapezoidal format into a CAD/CAM format, perform character and symbol recognition and support interactive editing. Formating times are still undetermined as the conversion routines for each CAD/CAM interface are still being developed. Character and symbol recognition takes approximately 20 minutes for 10,000 characters. Editing time will be dependent on the document itself factoring in such anomalies as line breaks, unrecognized characters, etc. ANA Tech's total processing time goal for an "E" size drawing is about one hour. Benchmarks performed for CAD/CAM systems will determine if this is feasible.

Broomall Industries

Broomall Industries recently announced and introduced several new products and enhancements to their automated data capture product line.

A new scanner with an active digitizing area 40" x 48" (document size up to 44" x 52") and with selectable 1000, 500, 250 and 200 lines per inch resolution was available for delivery in May, 1983.

An aperture card scanner with manual feed was available in June, an automatic card feeder in September, 1983.

A release of the raster to vector software including some symbol recognition features was made in the Spring 1983.

Intergraph

Intergraph has been making substantial R&D efforts toward providing an automated data capture product. These efforts are backed by a contract to deliver such a capabililty as part of a large interactive mapping system for the Australian Army.

At the ASP/ACSM Show in March, 1983, Intergraph was demonstrating their Scanned Data Capture System (SDCS). These capabilities were demonstrated on sample documents of contour and drainage map overlays. Character Recognition Processing was also demonstrated using a teach method of recognition.

This effort is the only known internal development going on within a major CAD/CAM system vendor since Computervision abandoned its efforts some years ago.

Interactive Systems

Interactive Systems introduced a vectorizing system with editing capabilities at ASP/ACSM 1982.

Basic vectorizing with some basic symbol recognition is available with promised character and symbol recognition at some future time.

Syscan (Kongsberg)

Syscan now offers raster editing which was developed in Europe. Systems combining raster and vector processing are now available for the following applications: vector processing for maps and mechanical drawings, raster processing of sketches and illustrations for technical manuals and publications.

Very basic symbol recognition is available to separate and eliminate text from a drawing and to recognize line widths.

Character recognition is in development and should be available in 1984.

Laser-Scan

Of the systems described in this article, Laser-Scan is the only one employing line following techniques as opposed to raster scanning and vectorizing. The overall process was described in the above-referenced articles. Since that time Laser-Scan has been setting up a U. S. subsidiary to begin a marketing thrust in the U. S. sometime during 1983. The new LASER-TRAK system was introduced at ASP/ACSM 1983.

The system demonstrated very efficient attribute tagging for map editing. It can provide raster scanning and vectorization and accept outputs of other data capture systems plus provide feature tagging of the data. Note that this is one of the prerequisites of good data base management, associativity.

The effectiveness of the Laser-Scan system was proved during the Falklands Islands war when a system was used to create the data needed to produce tactical operations maps within a couple of days.

Scitex

Scitex has made considerable progress since the preparation of the last article. This company combines raster data and vector data handling in the same system. Raster data is preprocessed before running vectorizing algorithms. This technique provides very clean raster data and lessens the decisions and iterations needed by vectorizing algorithms.

Post vectorizing processors are now available to do some steps of symbol recognition such as eliminating intermediate points on straight lines,, squaring corners of rectangles and recognizing ellipses.

Character recognition software is in development to handle machine produced lettering in any orientation. These developments include the "teach" method of recognizing the image.

On the hardware side, Scitex has recently introduced a new scanner/plotter which has been adapted from the laser plotter product. This provides a twofold thrust. First, it provides a large area monochrome gray scale scanner (40" x 72"). Second, it provides an overall system cost reduction from about $650,000. to about $500,000. since it eliminates one computer subsystem for plotting.

The goal of this system is to process a large size drawing in one hour. Scanning time on the 40" x 72" area is 20 minutes at 500 lines per inch.

Scitex will continue to offer its smaller scanner for color scanning applications.

REFERENCES

(1) "Automating Data Base Capture: Sanity or Folly" by Richard N. Stover. A series of five articles. Computer Graphics World, November '81; December '81; January '82; February '82 and March '82.

EXTENDED GRAPHICAL PLOTTING WITH THE PLANICOMP

Dierk Hobbie
Carl Zeiss, West Germany

ABSTRACT

To facilitate graphical plotting, two features have been added to the PLANICOMP system family:

- Computer-supported direct plotting and digital storage of the graphical data, and
- PLANICOMP integration in the workstation of the interactive graphics system made by INTERGRAPH.

Apart from the ZEISS DZ 7 Digital Tracing Table, various HP graphics terminals and plotters can be used for on-line digital plotting. Salient features of the "New Graphics" are, to list only a few, freely definable symbols, straight, circular or curved lines of different types inclusive of symbol lines and parallel lines, shaded areas and slope shadings. Storage of the graphical data allows plotting to be repeated any time also in a different location using other scales and orientations. An enhanced version of the PLANI-AS program allows the "New Graphics" to be used also with analog plotters.

A completely different capability is offered when the PLANICOMP is connected to an INTERGRAPH interactive graphics system. This system integration enables both real-time transfer of the points and lines measured with the PLANICOMP to the INTERGRAPH workstation, and optical superimposition of the INTERGRAPH screen with the PLANICOMP stereo image for direct visual checking.

MOSS--A STATE OF THE UNION ADDRESS

Carl N. Reed
Autometric Incorporated
Fort Collins, Colorado 80526 U.S.A.

ABSTRACT

In 1976, the U.S. Fish and Wildlife Service contracted for the development of a geographic information system. The product of this effort was a system called MOSS (Map Overlay and Statistical System). Since that time, MOSS has been adopted by the U.S. Fish and Wildlife Service and the U.S. Bureau of Land Management as their standard GIS. In addition, a number of other Federal Agencies and private companies have been using MOSS. This paper will begin with a discussion of the lessons learned in implementing a GIS and moving from the R&D realm into a production environment. This discussion will be followed by a summary of several representative projects. Finally, the current status of MOSS along with future enhancement will be described.

THE GEOGRAPHICALLY ENCODED STRING STRUCTURE (GESS) AS APPLIED TO THE 1980 CENSUS DIGITAL COUNTY BOUNDARY FILE

Roy Borgstede
Geographic Files Branch
Geography Division
U.S. Bureau of the Census
Washington, D.C. 20233 U.S.A.

ABSTRACT

Digital cartographic data bases have evolved over the past 2 decades from inflexible, highly structured, sequential files containing limited spatial information into the present flexible, more freely formed files containing spatial coordinates, as well as selected geographic relationships and information. The first file of this modern structure was the Geographic Base File/Dual Independent Map Encoding (GBF/DIME) files produced by the United States Bureau of the Census as a by-product of the address geocoding operations for the 1970 Decennial Census. A similar file structure was used to produce the 1970 DIME county boundary file (DIMECO). Now the Bureau has developed a file structure which is even more flexible than that used for the GBF/DIME while simultaneously carrying more geographic relationships and information than either the DIMECO or the GBF/DIME files. It is one which can be reduced easily to a simpler structure level if desired. The purpose of this paper is to describe the new geographically encoded string structure as it has been applied to the 1980 digital county boundary file.

APPLICATION OF A COMMERCIAL MAPPING/DATA MANAGEMENT SYSTEM TO FOREST LAND MANAGEMENT IN MAINE

Daniel E. Boss
Great Northern Paper
Millinocket, Maine 04462 U.S.A.

ABSTRACT

Great Northern Paper Company manages 2.4 million acres of forest land in northern Maine. In early 1980, an Intergraph minicomputer-based computer mapping system was acquired for the purpose of building a map-based information system. The key to GNP's use of the system is the plan to use map relationships, both for user-friendly retrieval of data base reports and to define data base relationships as required for different reporting purposes.

The unique land management problems which have been overcome all relate to the challenges of very complex polygon data with "island" conditions. They include:

1. The need for an efficient data entry system tailored to the entry of complex natural feature polygons and associated attribute data.

2. The need for data display methods which can handle large and complex polygons with "island" conditions to perform thematic patterning, dissolving and merging of reclassified polygon data, and separate display of areas with selected attributes.

3. The need for spatial relations processors which define new data base relationships based on spatial relationships in map elements which are linked to the data base.

The first two problems have been effectively solved with systems developed cooperatively by GNP and Intergraph during the past three years.

The last is currently being solved. It involves the definition of all classes of relationship between polygonal, lineal, and point type map features. GNP is solving the polygon overlay and intersection problem by using a grid type approach for the analysis of areal data, while maintaining polygonal representation for map update and storage. Processors to define relationships between attributes of points or lines and the attributes of containing polygons are now being developed. These developments will allow full relational data base retrieval from the graphically-linked data base, enabling GNP to meet the planned goals for its map-based information system.

EPPL: AN INNOVATIVE GRID GEOGRAPHIC INFORMATION SYSTEM LANGUAGE

Joseph Zastrow
Minnesota Land Management Information Center
The Minnesota Department of Energy, Planning and Development
St. Paul, MN 55101 U.S.A.

ABSTRACT

Grid or raster processing techniques are used by many environmental analysis systems for their cost effective capabilities. Minnesota has developed the Environmental Planning and Programming Language (EPPL) for use with its Land Management Information System. EPPL includes spatial analysis commands and multi-variable modeling techniques and contains several innovations that significantly enhance grid cell processing. The language is a translator, or "pre-compiler", which compiles user defined commands into a FORTRAN IV program optimized to only process the desired commands. EPPL data files combine run length and raster storage techniques which optimize storage, affording a 60% reduction in the disk space needed to maintain the Minnesota data base. EPPL also has the capability to mix FORTRAN code or subroutines with the standard commands for maximum flexibility. EPPL is fully interfaced to all major GIS related software capabilities: polygon processing, image processing, and digital elevation modeling. The power of this language allows the Minnesota Land Management Information Center to respond to a variety of requests for environmental data in an efficient manner.

A THREE DIMENSIONAL MODEL FOR EVALUATING POTENTIAL HYDRO-GEOTHERMAL RESOURCE UTILIZATION

Thomas W. Hodler
Department of Geography
The University of Georgia
Athens, Georgia 30602

ABSTRACT

A three dimensional surface representing the potential site specific utilization of hydro-geothermal resources was created using the Calcomp THREED program in conjunction with a weighted multiplication model. The four site variables incorporated into the model are temperature of the supplied water, topographic slope, land elevation, and cost of the water from source to site. These variables reflect the physical and economic constraints in the utilization of low temperature hydro-geothermal energy. The ordinally scaled model was created for an area in southcentral Oregon and northeastern California. The variables of slope, elevation, and temperature decay over transport distance were determined from USGS topographic maps. On the basis of the values of these variables and the distance of transport, a supply or pumping cost was then calculated. A 93 by 82 cell grid (each cell representing a four square kilometer area) was overlaid on the topographic maps. By multiplying the ordinally scaled assigned value for each variable (five values for temperature and three for each of the other variables), a maximum potential utilization value of 136 (0 to 135) was possible for each cell. The map of the individual model components and the composite map of maximum potential utilization values were generated by the THREED program. Of the 7,626 grid cells, slightly more than 5,000 cells lie within the area in which 70°C or hotter water can be supplied. Thus, two-thirds of the study area could be supplied with water that is hot enough (≥70°C) to be utilized for space heating. However, less than one percent of the grid cells generated the maximum possible value, i.e., the locations with the greatest potential for utilizing the resource. Ten percent of the area produced a composite value median to all variables, or a value of 24. The maps and data sets generated allow for both visual and numerical evaluation of the site specific potential for utilizing hydro-geothermal resources. Areas of greatest potential are most easily identified and lie in close proximity to the water source. Areas of moderate potential can be identified from the maps and may be identified for further, larger scale evaluation. The weighted multiplication model can be applied using the computer-assisted mapping package in order to evaluate the potential usage of the hydro-geothermal resource.

SOME PROBLEMS IN THE MAPPING OF GEOLOGICAL SURFACES

Christopher M. Gold
Earth Sciences Division
Alberta Environment, Canada

ABSTRACT

Most geological surfaces are sampled from drill hole or outcrop observations, and frequently multiple surfaces are recorded in the course of the same project. To complicate matters, information for all desired surfaces is frequently not available at every site.

Despite the development of more sophisticated techniques (local patches, kriging) simple weighted average techniques are still the most widely used. Perhaps one reason is that geological data sampling is notoriously ill-distributed and rarely samples a surface well enough to delineate even the wavelengths of interest. Consequently the methods used for CAD are rarely appropriate. Kriging methods, while in fact of weighted average type, require more manual intervention than is usually feasible. The worst problem of weighted average methods, that surfaces fail to pass precisely through the data points, is serious for oil industry users but can be eliminated by modifying the traditional rectangular-grid sampling scheme. A second problem, the selection of a suitable set of neighbouring data points to have non-zero weightings, can be handled with an appropriate data structure (e.g. triangulation). Most schemes for calculating the individual weightings are, however, rather arbitrary. However the set of non-zero weightings is achieved, the result is a weighting to be applied to each surface estimate derived from a neighbouring data point. This estimate may be the data point elevation, a projected value using slope estimates at each data point, or even a discrete value such as rock type.

Once the set of weights and surface estimates has been calculated, normally a weighted mean is the only value calculated. It is prefectly feasible, however, to estimate other statistical parameters, such as variance, from the same information. In addition, any particular contributing data point can be eliminated if it is beyond some specified boundary, falls outside the permissible range set by the under and overlying surfaces, etc.

Given the availability of this kind of information many useful operations may be performed. Error estimates can be made, either at data points themselves or at intermediate locations. Variance maps may indicate surface discontinuities or invalid data. And, finally, the neighbour selection and weighting function procedures themselves may be evaluated in terms of their contribution to overall map variance. The opportunity clearly exists for evaluating even these elementary techniques in an easily comprehensible and systematic fashion.

THEMATIC MAPPING

Orthogonal Three-Dimensional Views for Thematic Mapping 416
Kurt E. Brassel and Zissis Kiriakakis

Automatic Colouring of Maps According to the Elevation 426
B. Falcidieno, C. Gambaro and P. Sinigaglia

A Procedure for Shading Class Interval Regions 435
Robert Cromley

Accuracy of Viewing Three-Dimensional Bars in Perspective: The Case of Computer-Generated Pillars as a Thematic Map Symbolism Type (Abstract) 443
John Smyrnew

A Program for Automatic Name Placement 444
J. Ahn and H. Freeman

The Making of the Far Eastern Economic Review Economic and Social Atlas of the Pacific Basin 454
D.K. Forbes, C.C. Kissling, K. Mitchell, M.J. Taylor, N.J. Thrift and S. Wild

Towards an Electronic Atlas 464
Ewa Siekierska

Application of a Model of Dynamic Cartography to the Study of the Evolution of Population Density in Spain from 1900 to 1981 475
J.A. Cebrián de Miguel

Le traçage automatisé en mode négatif: la couche à tracer et la couche pelliculable 484
Denis Genest

Computer to Map: An Exercise in Communication (Abstract) 491
Rosemary Ommer and Cliff Wood

ORTHOGONAL THREE-DIMENSIONAL VIEWS FOR THEMATIC MAPPING

Kurt E. Brassel and Zissis Kiriakakis
Geographisches Institut
Universität Zürich-Irchel
Winterthurerstrasse 190
8057 Zürich

ABSTRACT

A new mapping method for statistical surfaces is proposed. It combines some properties of three-dimensional diagrams, oblique hill-shading and choropleth mapping. The model modifies the geometry of a discrete statistical surface consisting of a set of prisms. The slopes of the prisms are modestly tilted, and the new surface is viewed from a vertical viewpoint. Visual effectiveness is improved by using oblique hill-shading. The geometry and construction algorithms are explained, in addition some application problems are discussed.

1. INTRODUCTION

In a time when an abundance of spatial census information is available in digital form, computer mapping should make available adequate means for an efficient representation of such data. In particular, we are interested in methods that allow for the simultaneous display of multiple variables. The traditional means of displaying quantitative area information are the choropleth and graduated symbol techniques (Robinson, Sale and Morrison 1978). Various software packages have been developed to produce maps of these types (Brassel and Wasilenko 1980). Other methods include the three-dimensional portrayal of surfaces (Franklin 1979, Tobler 1975), the use of bars or pillars (Douglas 1975) or non-contiguous area cartograms (Olson 1976). In some test examples oblique hill-shading has been used to portray statistical surfaces (Brassel 1974, Brassel and Utano 1979).

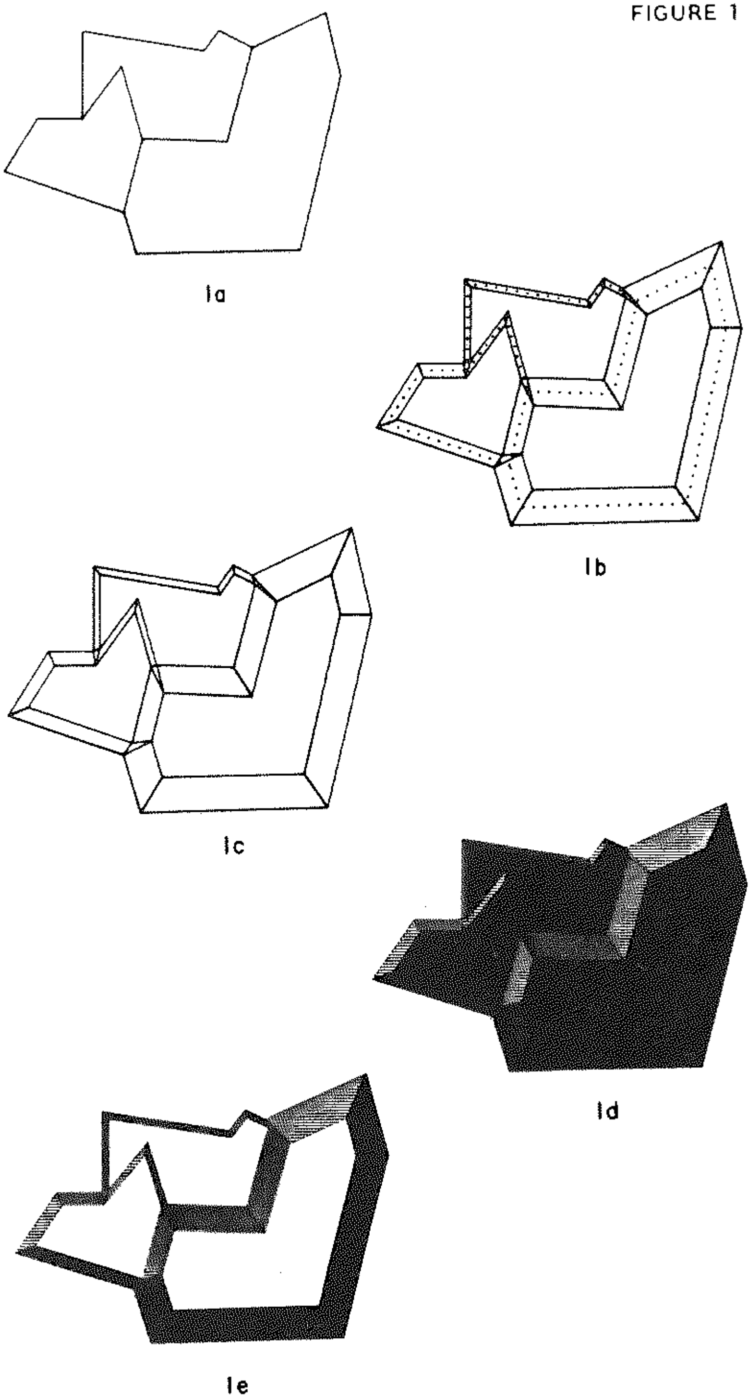
The present paper presents a mapping model which combines the concepts of the choropleth and the three-dimensional diagram techniques with automatic hill-shading. Whereas choropleth mapping varies the symbol density to express quantity, 3-d mapping uses hidden-feature and (pseudo-) perspective effects to suggest elevation, and hill-shading makes use of the perception properties of the illuminated surface.

The present model combines the three methods to map two variables simultaneously: the orthogonal three-dimensional view geometry in combination with hill-shading allows for the simulation of height of discrete regions, and choropleth shading of the polygon areas can then possibly be used to display a second variable.

Figure 1 illustrates the basic concept of orthogonal three-dimensional mapping: a discrete statistical surface in the form of a polygonal network structure (Figure 1a) is modified such that the polygons are separated by tilted planar slopes (1b, 1c) so that three-dimensional effects can be achieved by using oblique hill shading methods (1d). If the horizontal portions of the model are not shaded (1e), area shading symbolism may be applied to those polygons to portray some additional characteristics of the regions.

In this paper we discuss the basic concepts of the method (section 2), present a simplified algorithmic solution (section 3), and point out some geometrical and perception problems (section 4). Some comments on applications and unsolved problem areas conclude the presentation (section 5).

FIGURE 1



2. THE BASIC GEOMETRY

Assume a discrete statistical surface of geographical entities: a set of homogeneous regions is defined in the plane where a numerical value is assigned to each region. This value is considered the height of the region so that a discrete statistical surface of n regions can be considered a set of n prisms of individual height (Figure 2). The individual prisms are delimited by a pair of congruent polygons (horizontal) and a set of m vertical rectangles, where m is the number of discrete segments delimiting the region.

The basic idea of the orthogonal three-dimensional view concept is to modify the prisms by tilting all the side faces by a certain angle (Figure 3). As a general rule the tilting angle is kept constant throughout the entire statistical surface. Furthermore, the sides are tilted about a horizontal axis located at medium height of each side face. The geometry of the resulting surface is a function of the geometry of the original prisms, the tilting angle and the elevation difference between adjacent prisms. Figure 4 illustrates the most simple case where a single prism is surrounded by a planar surface. The model results in a polygon of reduced size plus a set of m trapezoids (Figure 4b). The m trapezoids are delimited sideways by the bisectors

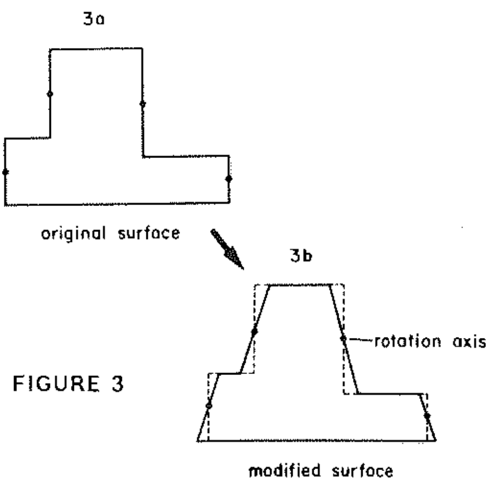


FIGURE 3

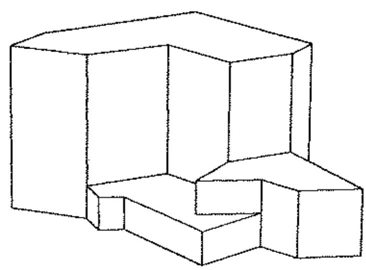
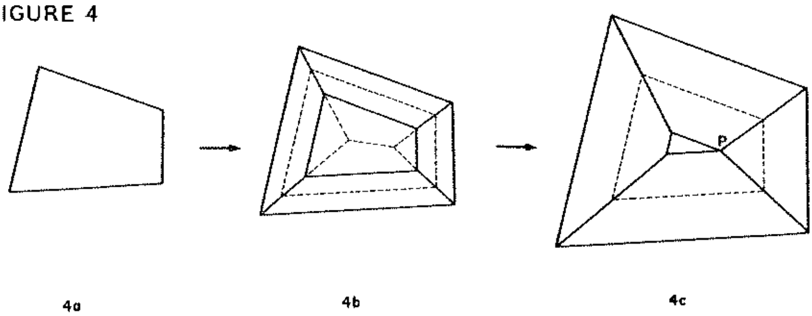


FIGURE 2
(Program CENVUE, Tobler 1975)

FIGURE 4



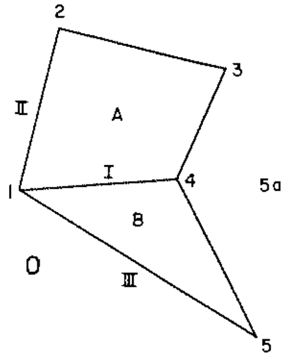
of adjacent polygon segments, and they are all of constant height. The height of the trapezoids is a function of the tilting angle and the elevation difference between the prism and the reference plane. Figure 4c shows that by increasing of tilt and/or elevation the top of the prism surface is progressively eroded up to a point where one or more edges of the original polygon collapse to a point P. The perpendicular distance of point P to the three adjacent edges of the original polygon is called the *critical distance* of the statistical surface configuration.

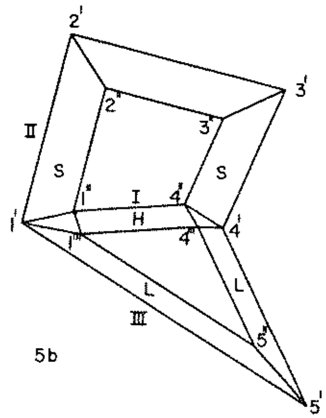
Figure 5a illustrates a statistical surface with two polygons A and B and a surrounding reference plane (O). The polygons are delimited by a set of points (1 through 5) which define six outline segments (1-2, 2-3 etc.). Points 1 and 4 are called nodes; these nodes are connected by three point chains called arcs (I: 14, II:1234, III: 154). The polygons are assigned non-identical elevation values (e.g. A=2, B=1, O=0). Elevation differences are constant along arcs (I:1, II:2, III:1). Figure 5b shows the model surface using constant tilting angles. The height of the trapezoids along arc II is double the height along arcs I and III. The points internal to arcs (2, 3, 5) are displaced along the bisector defined by the two adjacent segments to create points 2', 2''; 3', 3''; and 5', 5''. The two nodes (1,4) of the polygon network are represented in the final configuration by three points (1', 1'', 1'''; 4', 4'', 4''') where the relative displacement is a function of the ratio of elevation differences between the pairs of adjacent arcs. Again, by increasing the tilt a first polygon segment (1''-4'') collapses into a single point P (Figure 5c). Point P in Figure 5c is equidistant to the segments 12 and 34, but only at half the distance to segment 14.

Define $\Delta h(i)$ as the elevatipn difference pertaining to arc i and all its segments, then the relative distance *rdist* of point P to its neighbor segments is defined as $rdist(P) = dist(i) / \Delta h(i)$, where $dist(i)$ is the measured distance of point P to the segments of arc i. If point P is the point where the first segment of the entire model is collapsing, then $rdist(P)$ is called the *critical relative distance* of the configuration. Implicitly it is a measure for the *critical angle,* i.e. the maximum angle by which the vertical prism faces can be tilted without changing the topology of the statistical surface.

A topological node by definition is a point where three or more arcs converge. In a discrete statistical surface a general node separates horizontal planes on three (or more) levels. In the case where three arcs meet in a node, three sloping planes have to be constructed about that node: the first is separating the top level from the medium level plane: we call it here the *high* separating slope (H in Figure 5c); the second separates the medium level from the bottom level plane: we call it the *low* separating slope; and the third connects the bottom plane directly to the top plane: we call it the *spanning* (S) slope. In Figures 5b and 5c the spanning (S) slope is intersecting the high (H) and low (L) slopes. This is, however, not a general rule. The geometrical relationship between the three sloping planes is complicated and depends on both angular and altitude relations at a node. Figure 6 presents a case where the high slope separates the low and spanning slopes, Figure 7 gives an example where the low slope interrupts the high and spanning slopes. The situation is further complicated where more than three arcs meet at a single node.

The complex relationships between the slope planes imply a visual complexity of the resulting image. For this reason it is desirable to develop a model with a simiplified geometry.





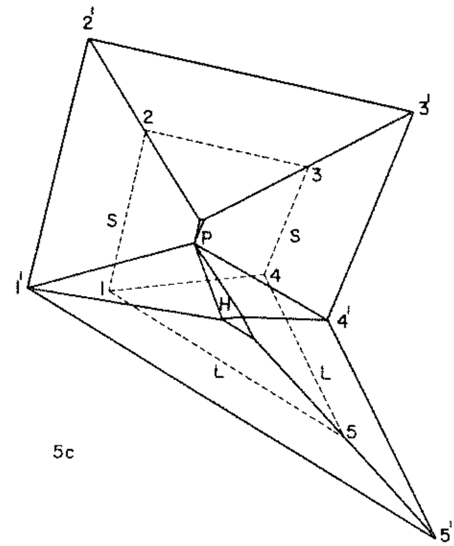


FIGURE 5

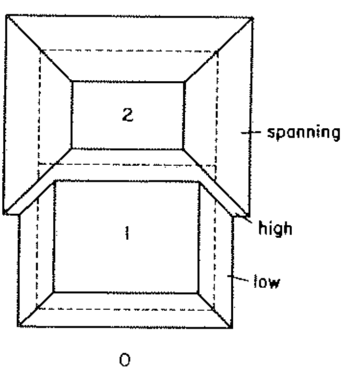


FIGURE 6

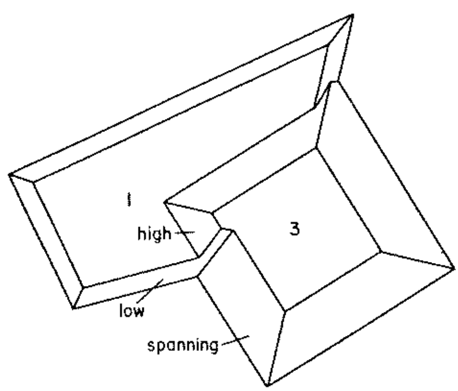


FIGURE 7

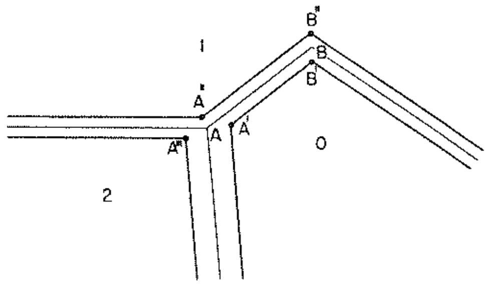
3. A SIMPLIFIED SOLUTION

We present here a simplified model of orthogonal three-dimensional mapping. Feature manipulation in two-dimensional space is sufficient for the construction process of this model.

Assuming that we know the critical angle of the overall statistical surface, we first select a global sloping angle equal to or smaller than the critical angle. Given a constant slope throughout the configuration, we know that the width of the sloping faces is proportional to the elevation difference along an arc. Based on this assumption we first construct a pair of parallel lines for each line segment, where the distance between segments and parallels reflects the elevation difference along the segment (Figure 8). The intersection of pairs of parallels of adjacent segments results in two *corner points* of the model surface. Near nodes three corner points are generated (A: A', A'', A'''), near cartographic points two corner points result (B: B', B''). The line B'B'' is a delimiter between two slope faces associated with two polygon segments, and the triangle A'A''A''' is considered a valid surface element of our simplified model. We associate a trapezoid (e.g. A'A''B''B') to all segments of the original network plus a triangle (e.g. A'A''A''') to all nodes in the system. The original two-dimensional polygon network (Figure 1a) is thus converted into a polygon network consisting of trapezoids, triangles and polygons (Figure 1c).

The model generation is organized as a sequential procedure. This procedure uses a topological polygon description file as input and creates a sequential file listing of the newly created trapezoids, triangles and polygons. In a subsequent step this output file is then displayed by a line plot or area shading program.

FIGURE 8



The processing sequence can be summarized as follows:

A. Select a first polygon
   - A1. Select a first arc.
     Compute and output a trapezoid for each arc segment.
     Compute and output a triangle pertaining to the end node of this arc.
   - A2. Process all remaining arcs in clockwise order (analog to A1).
   - A3. Compute and output the polygon (a modified version of the input polygon).

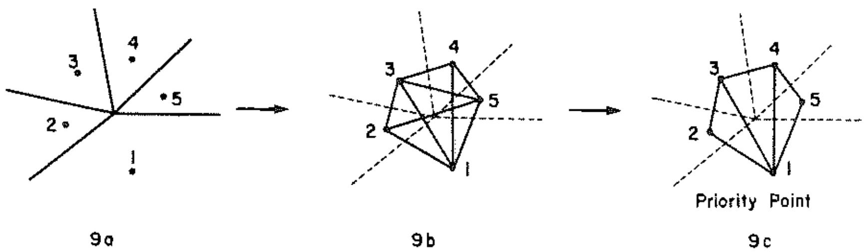
B. Process all remaining polygons analog to A, however:
   - do not process arcs which separate the active polygon from a polygon which has already been processed.
   - do not process nodes which pertain to a polygon which has already been processed.

C. Display the trapezoids, triangles and polygons by a line plot or area shading program.

The above general procedure has been amplified to adequately handle special cases.

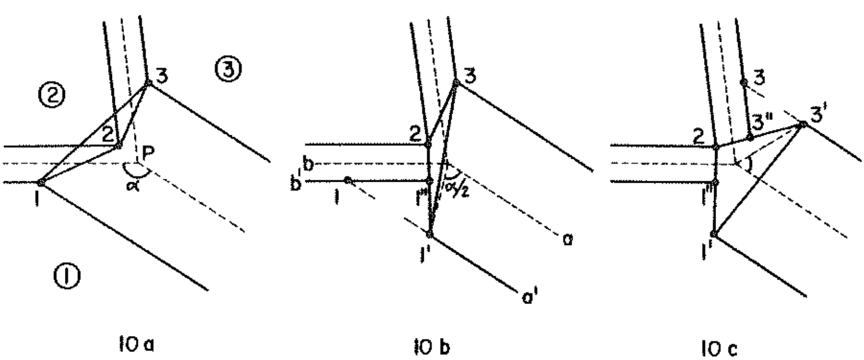
A first comment is needed for nodes where more than three arcs meet. In Figure 9a, for example, five polygons meet in a single node. Since inside each of the five polygons a corner point is computed, the set of five points could be used to create 6 triangles which partly overlap (Figure 9b). To avoid overlap, one of the corner points is declared a *priority point* and three triangles are then created where the priority point is always a triangle corner (Figure 9c). The priority point should be located at a 'central position' relative to all corner points, so that preferably the cyclical order of the corner points as seen from the priority point be identical to the cyclical order of the polygons themselves.

FIGURE 9



A second problem arises when the relative position of the three corner points to be connected is such that the triangle is 'flipped around', i.e. the cyclical order of the corner points is opposite to the cyclical order of the three polygons themselves: in Figure 10a the polygons 1, 2 and 3 are arranged in clockwise order about the common node P, the triangle of corner points 123, however, are in counter-clockwise order. The situation is handled as follows:

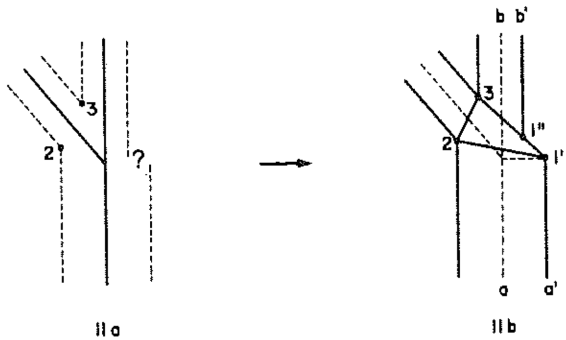
FIGURE 10



A1. Replace point 1 by a point 1' where 1' is at the intersection of the bisector of angle   and a' (the displaced segment pertaining to the arc with the larger elevation difference).
A2. Compute the triangle 1'23.
A3. Compute a point 1'' at the intersection of line 1'2 and b' (the displaced segment pertaining to the arc with the smaller elevation difference). Use this point 1'' for the description of the trapezoid pertaining to segment b.
A4. If the replacement of point 1 by the two points 1' and 1'' does not remove the reversal of the triangle, repeat the procedure by replacing point 3 in analog fashion by points 3' and 3'' (Figure 10c).

A third special case arises in a situation where two colinear arcs meet at a node (Figure 11a). Then two of the segment parallels are parallel, and one of the corner points is undefined. The solution is identical to the above problem: Compute the line perpendicular to the colinear arcs (which is the bisector of the straight angle) and find the intersection 1' of this bisector with the segment parallel a'. Compute the triangle 1'23 and find the point 1'' by intersecting 1'3 and b'. This point will be used to define the trapezoid pertaining to segment b.

FIGURE 11



4. THE CRITICAL ANGLE PROBLEM

We have introduced a simplified solution to the orthogonal three-dimensional view model which in certain instances is not geometrically correct. However, these simplifications in general do not negatively affect the visual perception of the statistical surface. Of greater importance is the fact that the critical angle may be very small; this depends on the configuration of the original polygonal network. The problem is that the sloping faces are very narrow and visually ineffective.

The critical angle is implicitly defined through the geometry of the polygonal network in conjunction with the elevation values assigned to the polygons. The critical angle can be increased only by a generalization of the polygon network. In general, the sizes of the minimum length segments have to be increased, narrow polygon gauges have to be eliminated, small angles have to be increased, etc. The orthogonal three-dimensional view method is effective only for statistical surfaces where the point density is low as related to the representation scale. Systematic studies will have to be conducted to determine the practical limitations of the method.

A further question relates to the method for finding the critical angle for a specific statistical surface. At this time the following algorithm is used:

A. Select the polygon of minimum area (*Pmin*); area: *amin*.
   A1. Set *rdist = SQRT(amin)/2 * h*, where *h* is equal to the minimum elevation difference pertaining to all the arcs of the polygon.
   A2. Construct a new "inside polygon" for polygon *Pmin* as follows: Replace each segment i by a segment i' parallel to it which is located at a distance *dist = rdist * h(i)* on the inside of the polygon. Segment i' is delimited by the parallels to the previous and following segments.
   A3. If the "inside polygon" is not self-crossing, set the preliminary relative critical distance *cdist = rdist* and initiate step B.
   A4. If the "inside polygon" is self-crossing, steps A2 and A3 are repetitively executed using binary fractions of *rdist* (*0.5*rdist, 0.25*rdist, 0.75*rdist* etc.) to find (with a desired resolution) a maximum relative distance *cdist* for polygon *Pmin*.
B. Process all remaining polygons as follows:
   B1. Compute the inside polygon at a distance *dist = cdist * h(i)*
   B2. If the inside polygon is not self-crossing, do not take any further action. If the inside polygon is self-crossing, use a binary search procedure to find a new cdist as a binary fraction of the old *cdist* (at a desired resolution).
C. *cdist* is the critical relative distance and is used to compute the critical angle.

5. APPLICATIONS AND CONCLUSIONS

This paper discusses the basic geometry of orthogonal three-dimensional thematic mapping and presents the algorithmic approaches for the construction of the model. At this time we do not have a detailed knowledge about the applicability of this method. We need some experience relating to its perception properties, in particular with complex statistical surfaces, extreme elevation differences, etc. Further studies will also have to be conducted to evaluate the use of the method for the representation of multivariate information, where the three-dimensonal effect is used to represent the first variable, and additional spatial variables are expressed by shading the horizontal portions of the model. The applicability can only be evaluated by systematic cognition tests based on application examples. Further efforts are required to improve the algorithms of the construction processes. Finally, we should also take efforts to find generalization processes for polygonal networks in order to generate generalized data bases which are adequate for orthogonal three-dimensional mapping.

ACKNOWLEDGEMENTS

Drs. G. Dorigo and H. Kishimoto have reviewed this paper and contributed figure 2, Mr. A. Herzog has supported the data base generation, Ms. Mueller and Mr. Steinmann have contributed to the reproduction of text and figures. All these contributions are gratefully acknowledged.

REFERENCES

Brassel, K. 1974, Ein Modell zur automatischen Schraeglichtschattierung: *International Yearbook of Cartography*, Vol. 14, pp. 66-77

Brassel, K. and Utano, J.J. 1979, Mapping from an Automated Display System: *Professional Geographer*, Vol 31/2, pp. 191-200

Brassel, K. and Wasilenko, M. 1980, Cartography and Graphics: in Marble D.F. (ed), *Computer Software for Spatial Data Handling*, Vol. 3

Douglas, D.H. 1979, The Pillar Mapping Program: Two Common Tasks of the thematic Cartographer: *Mapping Software and Cartographic Data Bases*, pp. 57-62, Harvard University

Franklin, W.R. 1979, PRISM - A Prism Plotting Program: *Mapping Software and Cartographic Data Bases*, pp.75-79, Harvard University

Olson, J.M. 1976, Noncontiguous Area Cartograms: *Professional Geographer*, Vol. 18, pp. 371-380

Robinson, A. and Sale, R. and Morrison, J. 1978, *Elements of Cartography* 4e, Wiley + Sons, New York

Tobler, W. 1975, A Computer Program to Draw Perspective Views of Geographical Data, Manuscript Dept. of Geography, Univ. of Michigan

AUTOMATIC COLOURING OF MAPS ACCORDING TO THE ELEVATION

B. Falcidieno (*), C. Gambaro (**), P. Sinigaglia (**)
(*) Istituto per la Matematica Applicata del C.N.R.
(**) Istituto Matematico dell' Universita',
Via L. Alberti, 4 - 16132 Genova - Italy

ABSTRACT

A method for the automatic production of colour banded elevation display is presented. This technique consists of a graphical postprocessor which colours the map beginning from its contour map, and has been usefully employed in bathymetric map production both to highlight errors or anomalies (difference maps, masks), and to display the final maps.

INTRODUCTION

Arbitrary surfaces, such as terrain and other natural phenomena, are generally described as numerical data (coordinates and relationships of points, lines and areas). Different data models encode properties and use storage in different ways. Encoded surfaces are generally called Digital Terrain Models (DTM), different formulations of which have been proposed, developed and used. /2,3,5/

These models are increasingly used in place of traditional topographic maps for geological, hydrographic and other environmental analysis, as well as for the storage and production of topographic maps themselves. Mark /5/ surveys the different DTM's employed and discusses their computational aspects. Although the DTM choice is subordinate to different conditions such as the application, the data acquisition method, the number of data and the available software facilities, Mark points out that the most frequently used are: triangulated irregular networks and regular elevation grids which define the terrain surface at every point of its spatial domain.

The problem we consider in this paper is how to provide the user with efficient and effective displays of the DTM on raster devices, since the display of a DTM before and after its entry into the digital Cartographic Data Base is particularly important because it provides the user with a powerful and efficient tool for error detection and gives the experienced observer a feel for the real appearence of the terrain. Thus, the display system must be quick in map production, to keep up with the interactive necessities of display, handling and editing of the DTM, and it has to produce effective, accurate and immediately legible maps to give the operator means for errors, anomalies and special feature detection.

Users generally prefer contour maps to other representations because they are planimetrically true and record elevations accurately.

Graphically, however, they leave much to be desired. Without information on the contour line elevations one cannot easily discern if the surface is rising or falling in a given direction so the appearence of the surface is visually ambiguous.
Three dimensional projections partially resolve this problem by presenting oblique perspectives of the surfaces. But these representations of terrain are not always appropriate or accurate. In most of them the vertical scale is exagerated and it is not possible to obtain elevation measurements because the base plane is hidden and perspective transformations can alter the vertical scale. Moreover parts of the surface can be hidden by closer features.
In any case representation of a 3D perspective of a surface involves several computations such as perspective transformation, shading, hidden part removal etc., which add further processing overhead.
An alternative is offered by the colour banded elevation displays, which use point elevation to modulate a colour intensity function /2/. The advantage of this display include showing slope irregularities, and giving the experienced observer a feel for the real appearence of the surface, even if it remains a 2½ dimensional method of surface representation.
Here we present a new method for the automatic production of colour banded elevation displays on raster devices. It consists of a graphical postprocessor which operates on the displayed map represented by contours coloured according to the elevation.
The method works in three steps. The first identifies the different regions by automatically locating an internal point, the second fills them and the third modifies the look-up table, colouring regions according to the elevation or to the user's requests.

REGION FILLING ALGORITHM

The problem of filling a region starting from its border has been solved in different ways which can be divided in two classes./6/.
The methods which belong to the first class start from the description of the contour as a polygon and decide what part of the plane lies within the interior, based on the equation of the lines. These techniques have been called polygon or edge filling methods.
Methods from the second class project the contour on the discrete plane and then determine the interior by examining the values of the pixels. These techniques are called pixel based or region filling.
There is a notable difference between the filling of polygons and the filling of regions. In the first case the region is defined by the vertices of the polygon and do not make hypotheses on the initial value of the buffer. In the second case the region is defined by the values of the pixels in the refresh buffer.
For solving our problem we considered only methods of the second class, completely rejecting algorithms belonging to the first one, since geometric information of the contour lines stored in the data base do not allow correct identification of the area to be filled.
In our data base only a first approximation of the contours in open

or closed polygonal lines is stored, while the displayed contours are produced by a smoothing algorithm which is performed in a post processing step./1/.

The pixel based class can be furthermore subdivided according to the technique they utilize to decide if a point is internal to an area. Some use parity check while others use the criterion of connectivity/6/. The first method is based on the fact that a straight line intersects a closed curve (such as the contour of a region) an even number of times.

The second method (connectivity filling) presumes that an internal point (seed) is given in addition to the contour and all the points are found that can be reached without intersecting it.

Between the two classes of algorithms the connectivity based ones are the most useful because they are not affected by the complexity and the irregularity of the contours, even if the necessity of knowing an internal point does not always make them easy to apply.

Different types of connectivity algorithms have been developed. In our application we chose Shani's algorithm /7/ because it was superior to the others for speed and flexibility /4/.

This algorithm raises the problem of filling as a variant of traversing a graph constructed on the region (fig. 1).

The region is subdivided in regular subregions (the nodes of the graph) and each boundary between two regular regions corresponds to an arc of the graph. The method consists of traversing the graph in successive phases characterized by one direction, colouring the nodes and pushing the nodes left suspended in the current direction on the stack.

To avoid eventually returning to previously coloured nodes, the incident arcs on the node in the direction opposite to the current direction, excluding the so-called 'entrance arc', are blocked (this corresponds to a tracing of temporary borders) and memorized in a queue from which they will be removed only when the stack is empty.

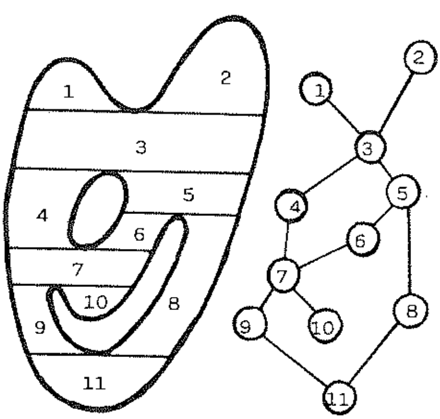


fig. 1

The stack and the queue are combined in a single structure called Queue-Stack from which it is possible to remove an element regardless of the position occupied, in addition to performing the operations of management of the stack and the queue.

The Shani's algorithm

A more precise description of the Shani's algorithm is given below.

```
Algorithm  FILL_REGION_1  (seed, filling_colour, boundary_colour)

// Where seed is an internal point to the region; filling_colour is
   the colour chosen for the filling; boundary_colour is the colour
   of the border. //

   let  direction be a possible direction
   SEARCH_ARC (seed, boundary_colour, p)
   BLOCK (p, boundary_colour)
   PUSH_ON_BOTTOM (p, -direction, Queue_Stack)
   PUSH_ON_TOP (p, direction, Queue_Stack)
   while  not  IS_EMPTY (Queue_Stack)  do
      POP(q, direction, Queue_Stack)
      if  q is blocked  then
         PAINT_ARC (q, filling_colour)
         SEARCH_PAINT_NODE (q, direction,filling_colour,boundary_colour)
         for  every arc r leaving NODE (q, direction)  do
            PUSH_ON_TOP (r, direction, Queue_Stack)
            PAINT_ARC (r, filling_colour)
         end  for
         for  every arc r leading to NODE (q, direction)  do
            BLOCK (r, boundary_colour)
            PUSH_ON_BOTTOM (r, - direction, Queue_Stack)
         end  for
         for  every blocked arc r leaving NODE (q, direction)  do
            REMOVE (r, Queue_Stack)
            PAINT_ARC (r, filling_colour)
         end  for
   end  while
end  FILL_REGION_1
```

In the description of the FILL_REGION_1 algorithm the following primitives have been used:

PUSH_ON_TOP (item1, item2, Queue_Stack)
which adds the elements item1 and item2 on the top of the Queue_Stack ;

PUSH_ON_BOTTOM (item1, item2, Queue_Stack)
which adds the elements item1 and item2 on the bottom of the Queue_Stack ;

POP (item1, item2, Queue_Stack)
which removes item1 and item2 from the top of the Queue_Stack ;

REMOVE (item1, item2, Queue_Stack)
which removes item1 and item2 from the Queue_Stack ;

IS EMPTY (Queue_Stack)
which returns the value TRUE if the Queue_Stack is empty, FALSE otherwise ;

SEARCH ARC (point, colour, arc)
which finds in the graph describing the region the arc bounded by the specified colour and containing the assigned point ;

PAINT_ARC (arc, colour)
which fills the arc by the assigned colour ;

BLOCK (arc, colour)
which blocks the specified arc painting it in colour ;

SEARCH_PAINT_NODE (arc, direction, filling_colour, boundary_colour)
which finds and paints in the filling_colour the node connected by arc in the assigned direction and bounded by a boundary of the specified colour ;

NODE (arc, direction)
which returns the node connected by arc in the assigned direction.

THE NEW ALGORITHM

The problem which remains to be solved is the automatic determination of the inputs necessary to the application of the connectivity algorithm to fill each region, that is the coordinates of an internal point (seed), the colours of the border and the filling colours.

Automatic determination of a seed

The displayed contour maps, which are the input of the algorithm,contain closed lines, open lines intersecting the border of the screen in two points, interrupted internal lines and interrupted lines intersecting the border of the screen in one point (fig. 2).

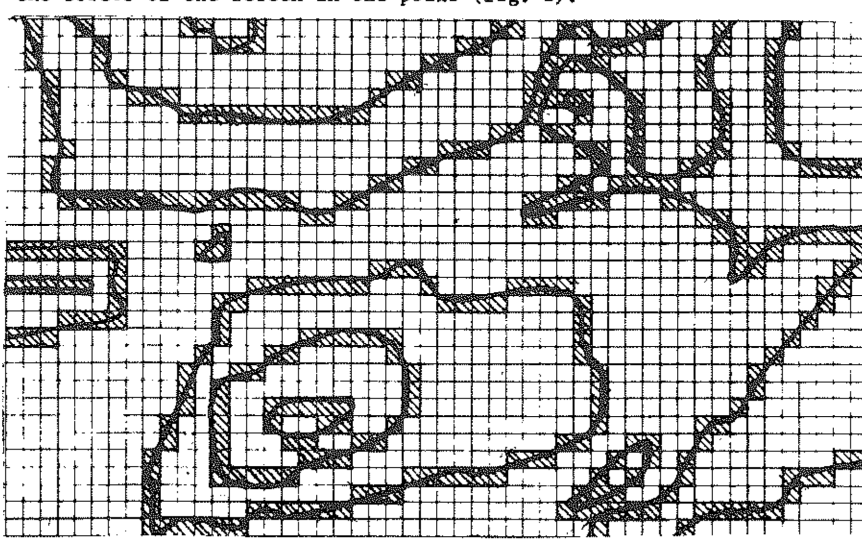


fig. 2

Since the scaling and clipping transformations and the graphic resolution produce new kinds of contour lines which are not always consistent with the characteristics of closure and non intersection that are peculiar of the definition of isolines, the existing algorithms for the automatic research of an internal point, which adopt a parity check criterion, must be rejected because of their lack of efficiency and reliability when applied to this particular type of regions.
The problem of finding the internal point in this situation where the region can be closed, open and with some remarkable irregularities in the border can be solved by drawing the contour lines on a given background colour which must be privileged and different from all other colours used both for the contour lines and for the filling of the regions present in the map.
This allows distinguishing all the areas to be filled since each time the algorithm identifies a pixel P having the background colour the area connected to it is automatically determined. Since the set of pixels connected to P and having the same colour represents its interior, while the set of pixels connected to P but having different colours represents its border.
Thus the map can be coloured by scanning all pixels of the screen and by filling the connected area for every pixel having the background colour.

Determination of the colours

Determining the filling and boundary colours is not based on a preprocessing step but it is effected during and after the filling process by taking advantage of the difference existing between logical and physical colour on Raster display and using the contour analysis, which is already effected by the filling algorithm. The logical colour is the value present in the refresh buffer, which points to an address in the Look Up Table (LUT) in which the physical colour is stored, and the physical colour is the value by which the intensity of the three primary colours (Red, Green, Blue) is determined.
For this reason the LUT must be constructed with reserved addresses for privileged colours: one for the background, N colours for displaying elevation, where N is a priori determined by the user and equal to the maximum number of elevation in the map, and M filling colours, where M is determined by the algorithm and equal to the number of regions simultaneously present in the displayed map.
A suitable logical colour is used to determine the filling colour whose physical colour is computed as the average of the physical colours intercepted in the border after the filling is finished.

The modified filling algorithm

The assumption of a uniform background simplifies Shani's algorithm because it avoids execution of the procedures BLOCK and REMOVE thus eliminating several searches in the Queue-Stack, while the possibility of different colours in the border and their successive storage introduce a generalization and further processing overhead.
The handling of the needed colours can be improved by further condi-

tions in the LUT organization, that is, giving the 0 address to the background colour and putting the colours of curves in a contiguous zone.
So checking if a pixel belongs to the contour is reduced to verifying if its colour is greater than zero and the identification of the colours of the border is done by marking the related array.
The modified filling algorithm consists of the following steps:

```
Algorithm FILL_REGION_2 (seed, filling_colour, boundary_colour)
   let direction be a possible direction
   SEARCH_ARC (seed, p, boundary_colour)
   PUSH_ON_BOTTOM (p, -direction, Queue_Stack)
   PUSH_ON_TOP (p, direction, Queue_Stack)
   while not IS_EMPTY (Queue_Stack) do
      POP (q, direction, Queue_Stack)
      if colour of q ≠ filling_colour then
         SEARCH_PAINT_NODE (q,direction,filling_colour,boundary_colour)
         for every arc r leaving NODE (q, direction) do
            PUSH_ON_TOP (r,direction, Queue_Stack)
         end for
         for every arc r leading to NODE (q, direction) do
            PUSH_ON_BOTTOM (r, -direction, Queue_Stack)
         end for
      end if
   end while
end FILL_REGION_2
```

In the FILL_REGION_2 algorithm two primitives are modified as follows:

SEARCH_ARC (point, arc, boundary_colour)
: which finds the arc of the region containing the assigned point and marks the encountered boundary colours in the corresponding array;

SEARCH_PAINT_NODE (arc, direction, filling_colour, boundary_colour)
: which finds and paints in the filling_colour the node connected by arc in the assigned direction. In addition, it marks the locations corresponding to the found boundary colours in the boundary colour array.

The general algorithm which produces the colour to elevation display of a map starting from its contour representation is described below:

```
Algorithm FILL_MAP
// boundary_colour array contains all the colours of the contours. //
   for every row on the screen do
      for every pixel p in the current row do
         if colour of p = background_colour then
            let current_colour be the first free position in the LUT
            FILL_REGION_2 (p, current_colour, boundary_colour)
            LUT (current_colour) ← AVERAGE_COLOUR (boundary_colour)
         end if
      end for
   end for
end FILL_MAP
```

In the FILL_MAP algorithm a new primitive is used:
AVERAGE_COLOUR (boundary_colour)
which returns the average colour of the marked colours in the boundary_colour array.

Figure 3 shows the contour map of a region where each line is coloured according to the elevation. Figure 4 is the corrisponding colour banded elevation display produced by the proposed method. (Both figures are shown here in black and white).

CONCLUDING REMARKS

This method which combines the advantages of both contour maps and colour banded elevation display, has been usefully employed in bathymetric map production, both to highlight irregularities or anomalies (difference maps, masks), and to display the final maps.

Since it is a graphical postprocessor, which operates on displayed maps represented by contour lines, it is independent of the particular DTM recorded in the data base and therefore it is generally applicable. Furthermore it is quick because it is completely automatic and its time complexity in the average case increases with the number of regions simultaneously displayed on the map.

ACKNOWLEDGMENTS

The reserch was supported by the Italian National Research Council (Progetto Finalizzato Informatica - Sottoprogetto P1 - Obbiettivo SofMat). The authors also gratefully acknowledge C. Pienovi for her helpful comments.

REFERENCES

/1/ De Floriani,L., Dettori,G., Falcidieno,B., Pienovi,C., Gianuzzi,V. 1981, A Graphical System for Geographic Data Processing in a Minicomputer environment: Signal Processing, Vol. 3, pp. 253-257

/2/ De Gree,M. 1982, Digital Elevation Model Image Display and Editing: Proceeding ACSM - ASP Fall convention, Ft Lauderdale, Sept. 22

/3/ Dutton,G. 1982, Land Alive: Perspectives in Computing, Vol.2 N.1

/4/ Gambaro,C., Sinigaglia,P. 1983, Contour and Region Filling Methods: Proceeding of the Second IASTED International Symposium on Robotics and Automation, Lugano (To appear)

/5/ Mark,D.M. 1978, Concepts of Data Structure for Digital Terrain Models: Proceedings of the DTM Symposium ASP-ACSM, St Louis

/6/ Pavlidis,T. 1982, Algorithms for Graphics and Image Processing: Springer Verlag

/7/ Shani,U. 1980, Filling Regions in Binary Raster Images: a Graph - teoretic approach: <u>Siggraph Proceeding</u>, pp. 321-327.

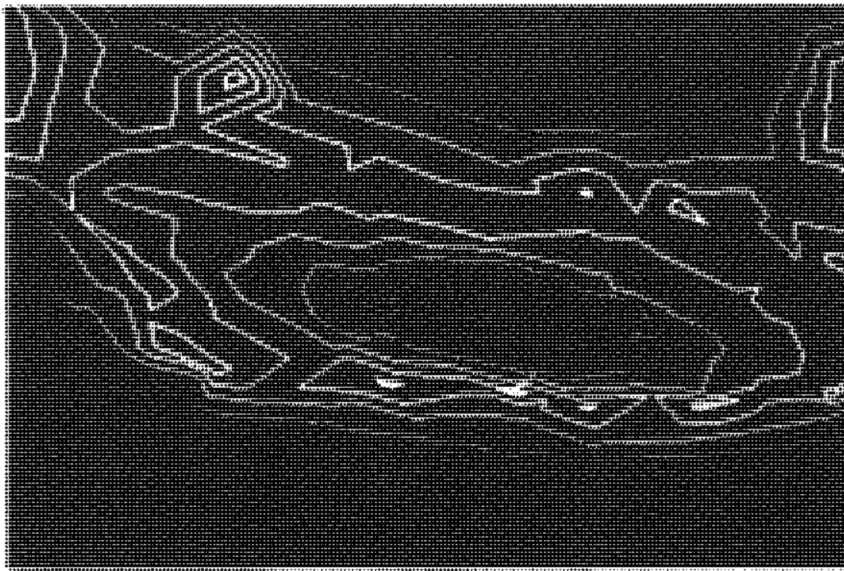


fig. 3

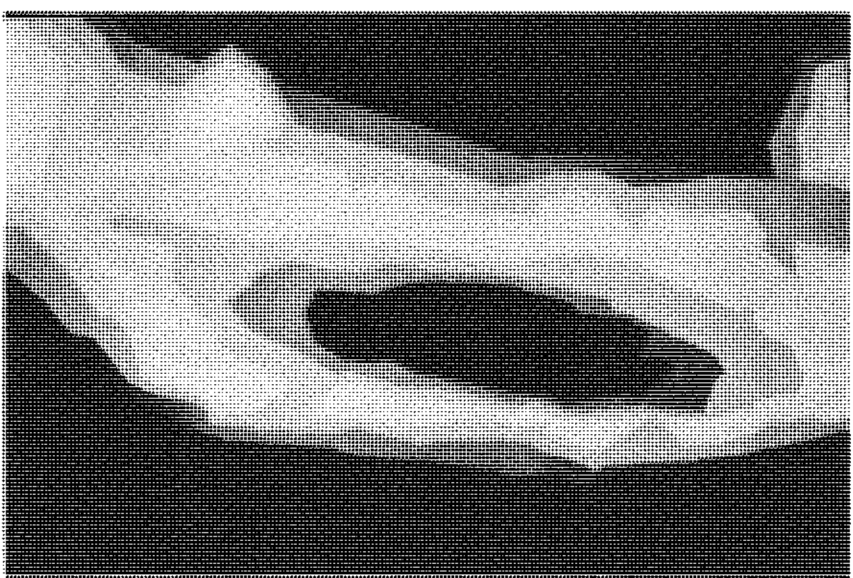


fig. 4

A PROCEDURE FOR SHADING CLASS INTERVAL REGIONS

Robert G. Cromley
University of Connecticut
Storrs, CT 06828

ABSTRACT

Although improvements have been made in the efficiency of polygon line-shading algorithms, choropleth maps are still drawn on vector-mode devices one data zone at a time. This paper presents an alternative method for constructing choropleth maps by shading composite class interval regions. By storing the base map in a triangular data structure (Cromley, 1983), it is possible to identify the boundaries of the class interval region using a procedure similar to one developed by Elfick (1979) for finding contour lines on an isarithmic map. Overall computational time is reduced by this method since a large number of internal boundaries are eliminated.

INTRODUCTION

Efforts to improve the computational efficiency of drawing choropleth maps on a vector-mode device have concentrated on improving the performance of polygon line-shading algorithms (Brassel and Fegeas, 1979; Lee, 1981; Cromley, (1982). The overall mapping strategy, however, has been to shade individual data zones in sequence. Thus, polygon line-shading algorithms must orient the family of shade lines associated with a particular shading pattern with respect to the origin so that the shade lines passing through two regions belonging to the same class interval and sharing a common boundary will exactly align.

A disadvantage of this approach is that an individual shade line is broken into a series of segments if it passes through more than one data zone. Shade line S in Figure 1 passes through regions A and B. If region A is shaded independently of region B, then S is drawn as three segments when shading region A and as four segments when shading region B. If the two regions were joined together and shaded as one, then S is drawn as one complete segment. If, for a given thematic map, the data values for regions A and B belong to the same class interval, then these regions can be united.

A second disadvantage is that all boundaries between data zones are processed twice as part of the outline of two different regions. This considerably increases the number of cartographic points that must be processed when shading individual regions. Cartographers have recognized the inefficiency of storing the individual outlines of respective data zones (Peucker and Chrisman, 1975) but these zones usually are processed in this manner for shading.

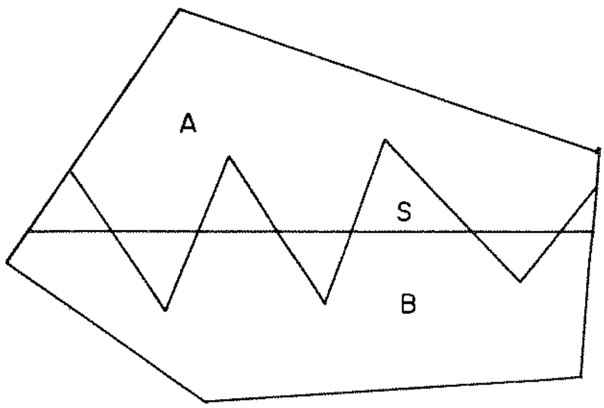


Figure 1. A typical shade line

This paper presents an alternative strategy for drawing choropleth maps by shading individual class interval regions. One class interval region would be the composite area of all individual data zones belonging to the same class interval. For a four interval map, then, only four regions would be shaded although each region may have many internal and external islands. The class interval regions are compiled using a triangular data structure for storing choropleth base maps (Cromley, 1983). This approach eliminates all shared boundaries between data zones of the same class interval reducing the total number of points that must be processed by polygon line-shading algorithms.

BACKGROUND

Choropleth maps are planar representations of three-dimensional stepped surfaces. The relevant elements of a choropleth base map are nodes, arcs, and regions (see Brassel, 1978). Nodes (N) are zero-dimensional locational identifiers formed by the intersection of three arcs. An arc (A) is a linear sequence of points that form the boundary between two regions. Regions (R) are two-dimensional planar surfaces for which data values are collected.

Of these elements, nodes are the most important topologically and form the basis for organizing the data structure. Each node is connected to three other nodes by three different arcs. Moving from a node to one of its adjoining nodes, one region cobounding the arc is defined as the right-hand region and the other as the left-hand region. In Figure 2, R7 is the right-hand region and R1 the left moving from node N2 to node N3. The adjoining nodes, connecting arcs and right-hand regions, define the neighborhood information for each node. Using this neighborhood information, a node reference file can be constructed similar to Elfick's (1979) triangulation structure for contour maps (Cromley, 1983).

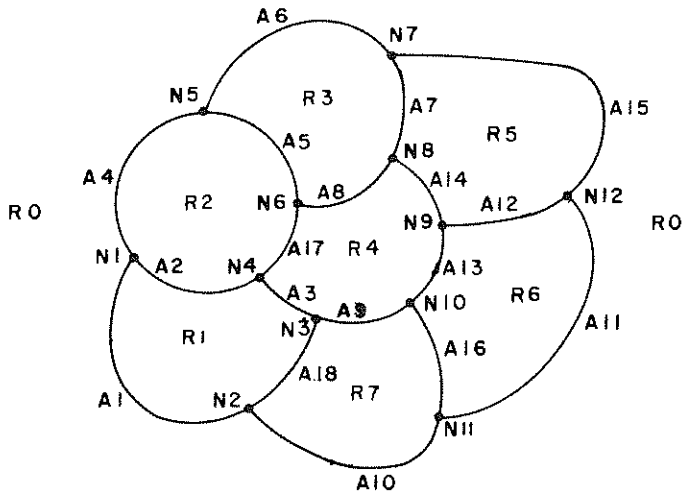


Figure 2. An example base map

TABLE 1
The Node Reference File

Reference Node	Adjoining Nodes			Right-Hand Regions			Connecting Arcs		
N1	N2	N4	N5	R0	R1	R2	A1	A2	A4
N2	N11	N3	N1	R0	R7	R1	A10	A18	A1
N3	N4	N2	N10	R4	R1	R7	A3	A18	A9
N4	N1	N3	N6	R2	R1	R4	A2	A3	A17
N5	N1	N6	N7	R0	R2	R3	A4	A5	A6
N6	N4	N8	N5	R2	R4	R3	A17	A8	A5
N7	N5	N8	N12	R0	R3	R5	A6	A7	A15
N8	N6	N9	N7	R3	R4	R5	A8	A14	A17
N9	N8	N10	N12	R5	R4	R6	A14	A13	A12
N10	N3	N11	N9	R4	R7	R6	A9	A16	A13
N11	N10	N2	N12	R6	R7	R0	A16	A10	A11
N12	N7	N9	N11	R0	R5	R6	A15	A12	A11

Table 1 is the node reference file for the base map presented in Figure 2. The first three entries of each record are the node numbers for the three adjoining nodes recorded in counterclockwise direction. The next three entries are the region numbers of the corresponding right-hand region between the current node and each adjoining node. The background is also a right-hand region when appropriate. In Table 1, it is denoted by R0. The last three entries in each record are the arc numbers for the corresponding arc connecting the current node to each of the adjoining nodes.

An arc file must also be established that contains a locus of points approximating the boundary between two regions. The starting node for each arc is included to determine the direction of each arc. Using these two files it is possible to enumerate class interval regions.

THE CLASS INTERVAL PROCEDURE

The outlines of individual data zones can be retrieved by cross-referencing the node reference file and arc file with a pointer file containing a starting node for each region (see Cromley, 1983). Once the entire outline of an individual data zone has been enumerated, it can be shaded in the tone for its class interval. This process, though, does not take advantage of the fact that many regions in the same class interval share common boundaries. Thus, all points and boundaries are processed by the shading routines. Because the choropleth base map has been stored in a triangulation data structure, a procedure analogous to the method defined by Elfick (1979) for constructing contour lines on an isarithmic map can be developed to composite class interval regions in which all interval boundaries have been deleted.

The problem then is to retrieve the outline of the perimeter of each class interval region. For each node in the base map, one of three situations can occur: 1) each of the three regions sharing that node belong to the same given class interval, 2) none of the regions belong to the given interval, or 3) some of the regions belong to the same interval while the remainder do not.

In the first case, the node lies in the interior of the interval region and none of the arcs emanating from it form part of the perimeter of the class interval region. In the second case, the node lies outside the interval region and again it is impossible for any of its connecting arcs to be part of the perimeter. Only in the third situation will the node lie on the perimeter of the interval region and two of the three connecting arcs will be part of the interval region perimeter.

Each adjoining node shares two regions with the node currently under examination. Once an arc associated with this current node is found to be part of the perimeter, one can move along it to its adjoining node. Only the unchecked region must be examined to determine if it lies inside or outside the class interval region. The following algorithm uses these facts to "walk" around the perimeter of the various class interval regions for a choropleth map having K nodes:

Step (1) Select the class interval; if all class interval regions have been enumerated, STOP; otherwise go to Step (2).

Step (2) Set the flag for each node to an "off" position; go to the next step.

Step (3) Set the node reference value, N, to 0; go to the next step.

Step (4) Increment N = N + 1; if N greater than K, go to Step (11); otherwise go to the next step.

Step (5) If the flag for node N is turned "on," go to Step (4); otherwise set M = N and go to the next step.

Step (6) If all neighbor regions of M belong to the current class interval, go to Step (4); if none of the neighbor regions belong to the current class interval, go to Step (4); otherwise go to the next step.

Step (7) If two regions belong to the current interval and one does not, go to Step (10); otherwise set the value of PR, the current perimeter region, to the region reference value of the only neighbor region of node M belonging to the current interval; go to the next step.

Step (8) Find M', the next clockwise perimeter node, in the node reference file as the adjoining node for PR; turn the flag for M' "on;" add the arc connecting M and M' to the set of perimeter arcs; if M' = N, go to Step (4); otherwise go to the next step.

Step (9) If the unchecked region of M', TR, belongs to the current class interval, set PR = TR and M = M' and go to Step (8); otherwise set M = M' and go to Step (8).

Step (10) Set the value of PR to be the region reference value of the right-hand region belonging to the class interval that corresponds to a perimeter arc of the class interval region; go to Step (8).

Step (11) Shade the composite class interval region; go to Step (1).

In this algorithm, the background region, R0, is always given a dummy value so that it does not belong to any class interval. Thus, all border arcs will be part of the perimeter for one of the class interval regions.

ALGORITHM IMPLEMENTATION

The above algorithm (CIP1) was encoded in FORTRAN IV and implemented on an IBM 370. CIP1 was time checked against the standard method for enumerating and shading individual data zones (STAND) using three test examples of the 120 counties of Kentucky. The first example is a "four-color" map where each arc in the base map was part of the perimeter for some class interval region (Figure 3a). In the second example (Figure 3b) each class interval had the same number of observations and there was slight spatial clustering of regions in each class interval. Finally, the last example (Figure 3c) had a skewed distribution for the number of observations in each interval and a high degree of spatial clustering within each interval.

The CIP1 code was most efficient in the latter two examples (see Table 2). The standard approach performed best when no points or arcs were omitted. The computational trade-offs between the two approaches for drawing a choropleth map center on the number of points associated with each region being processed. The standard method involves fewer points per individual and region but almost always more total points per map while the class interval approach has the opposite features.

The performance of the class interval approach would be improved if individual external islands within each class region could be isolated and shaded independently since shading algorithms are sensitive to the total number of points in an outline. The CIP1 code was modified to isolate external islands that are single data zones (code CIP2). This modified code did improve the efficiency of the class interval approach when fewer points and arcs were eliminated (see Table 2).

TABLE 2
Computational Time Checks

CIP1

	Number of Points	Percent of Total Points	Time (milliseconds)
Test Example 1	4348	99.6	14086
Test Example 2	3714	85.1	12223
Test Example 3	1556	35.6	5050

STAND

	Number of Points	Percent of Total Points	Time (milliseconds)
Test Example 1	4364	100.0	11599
Test Example 2	4364	100.0	12529
Test Example 3	4364	100.0	12699

CIP2

	Number of Points	Percent of Total Points	Time (milliseconds)
Test Example 1	4348	99.6	11469
Test Example 2	3714	85.1	12133
Test Example 3	1556	35.6	4836

CONCLUSION

The class interval approach represents an alternative method for drawing choropleth maps by machine. It rein-

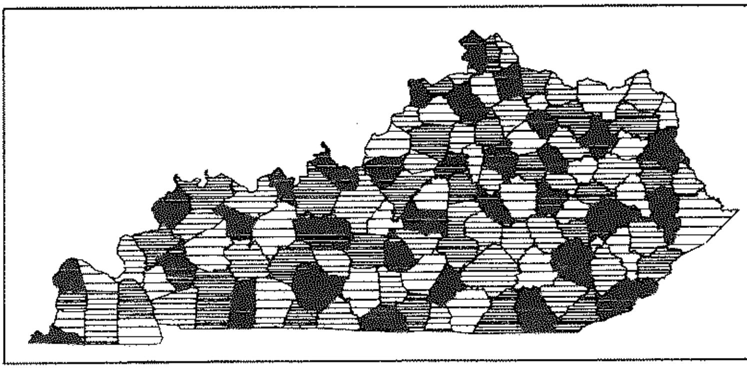


Figure 3a. The first test map

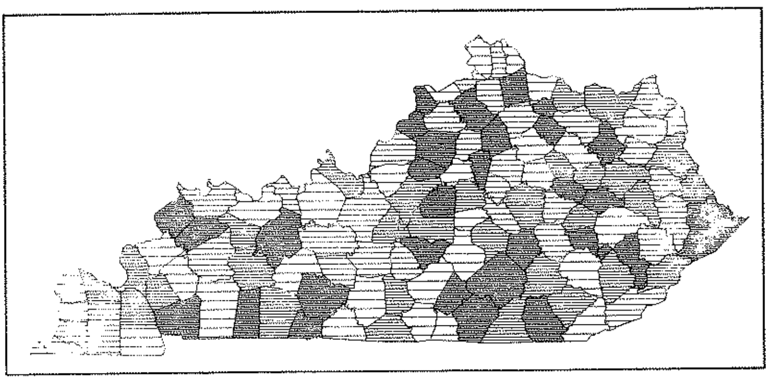


Figure 3b. The second test map

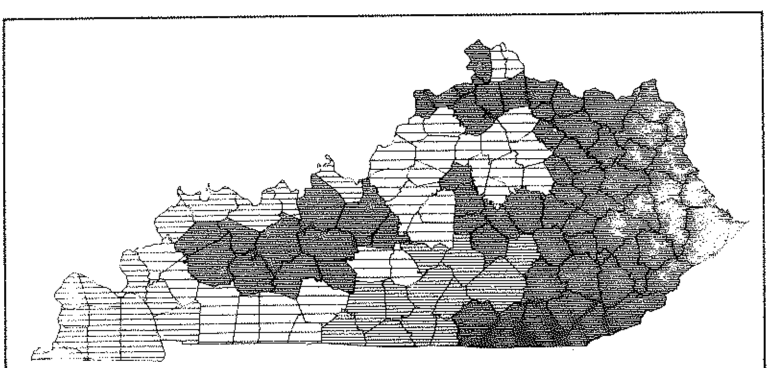


Figure 3c. The third test map

forces the general proposition that data structures based on the storage of individual data zones are inefficient and obsolete. The procedure presented here strongly parallels Elfick's (1979) triangular mesh method for drawing contour maps. Using different subroutines, a general procedure is possible, then, for the construction of all planar representations of three-dimensional surfaces. Topologically speaking, constructing a choropleth map is no different from constructing an isarithmic map.

ACKNOWLEDGMENTS

The author wishes to thank Gyula Pauer, Jr. for final preparation of all figures.

REFERENCES

Brassel, K. 1978, A Topological Data Structure for Multi-Element Map Processing: Harvard Papers on Geographical Information Systems

Brassel, K. and Fegeas, R. 1979, An Algorithm for Shading of Regions on Vector Display Devices: Proceedings, SIGGRAPH, Vol. 13, pp. 126-133.

Cromley, R. 1982, The 'Peak,Pit,Pass' Procedure for Polygon Line-Shading, Paper presented at the Annual Meeting of the Association of American Geographers, San Antonio, Texas.

Cromley, R. 1983, A Triangulation Data Structure for Choropleth Base Map, submitted to Cartographic Journal.

Elfick, M. H. 1979, Contouring by Use of a Triangular Mesh: Cartographic Journal, Vol. 16, pp. 24-29.

Lee, D. T. 1981, Shading of Regions on Vector Display Devices: Computer Graphics, Vol. 15, pp. 37-44.

Peucker, T. and Chrisman, N. 1975, Cartographic Data Structures: The American Cartographer, Vol. 2, pp. 55-69.

ACCURACY OF VIEWING THREE DIMENSIONAL BARS IN PERSPECTIVE: THE CASE OF COMPUTER GENERATED PILLARS AS A THEMATIC MAP SYMBOLISM TYPE

John M. Smyrnew

Office of Institutional Studies
Athabaska University
Edmonton Alberta
Canada

ABSTRACT

Today, in the business community, scientists, and the public in general, are trying to deal effectively with the information explosion brought about through the miracles of the chip. How is this data explosion to be handled so that the information becomes available and comprehensible to everyone? An answer lies in finding effective methods to communicate statistics quickly, accurately, and more intelligibly. One aspect of this problem is found in the symbology used in data presentation. In this paper, a closer look is taken at the three-dimensional bar, or pillar, as a symbol for displaying differences and similarities of quantities as applied to geographic distributions. Specifically, the map form analysed is the the computer generated pillar map, which displays a geographic distribution by a perspective view of a map surface on which square based pillars which vary in height stand. The base area of all pillars remains constant, therefore the value for each geographic point is symbolized by the apparent height of the pillar. The functional usefulness of this symbolism is dependent on the ability of map readers to compare and classify it around the quantities it represents. This representation must be based on precise knowledge of the relationship between the capabilities and the perception of the use, as well as the characteristics of a symbol that is incorporated into a map. A method is described that was developed to provide information that could be used towards more effective presentation of the pillar map form. This method used survey and statistical measurement techniques. Summary results and suggestions for more potent uses of the pillar symbolism in both map and non-map environments are given. Such knowledge is fundamental in developing principles of effective cartographic communication that may be employed in creating maps.

A PROGRAM FOR AUTOMATIC NAME PLACEMENT

J. Ahn and H. Freeman
Image Processing Laboratory
Rensselaer Polytechnic Institute
Troy, NY 12181

ABSTRACT

This paper describes the development of a computer program that will automatically place area, line, and point feature names on a map. The program follows established cartographic rules, such as having area names show the extent and orientation of an area feature and having a line feature name follow the curvature of the feature. The program achieves unambiguous placement of point feature names, prevents overlap between names and point features or other names, and strives for aesthetically pleasing overall effects. Input data consists of the map description and associated gazeteer. The program is able to accommodate different fonts, character sizes, character spacing, and colors.

INTRODUCTION

Feature name labels form an important part of a map for several reasons. The most obvious reason is that the labels tell us the names of the geographical entities represented. However, there are additional reasons. A label can give us an implicit understanding of the linear or area extent and orientation of a map feature, of the nominal class to which the geographical feature belongs, and of the relative size or importance of the feature. This is accomplished through the font, size, spacing and placement of the label. Finally, feature name labels serve also to enhance the aesthetic appearance of a map. It has been said that "good form and placing of type makes a good map" [Imhof 1975].

The quality of a map depends strongly on the way it is annotated. However, despite all the advances made in map data processing in recent years, the annotation or name placement process has not been automated to any significant degree. This paper describes the development of a computer program which will automatically place names on a map -- for point, line and area features -- applying the rules used by cartographers. The process is to be fully automatic; though, an interactive editing system will be provided to make minor name placement adjustments or corrections in special cases. [Hirsch 1980, Kelly 1980]

Basic Definitions

The data represented in a map is usually of one of three types: point data, line data, and area data. Point data refer to small, localized places on a map. Examples of point data are cities, villages, mountain peaks, ports, towns, mines, churches, and historic sites. For purposes of computer processing, point data can be represented by x,y-coordinate pairs.

Line data (more precisely, curve data) consists of geographical features that have linear or ribbon-like extent. Examples of line data include rivers, highways, canals, railroads, streets and ship courses. Line data can be described in a computer by the x,y-coordinate values of one or both endpoint(s) and a description of the curve that connects the endpoints. Of the several methods available for these purposes, one is use of a list of closely-spaced point coordinates along the curve; another is chain coding [Freeman 1974].

The third type of data on a map is area data, which represents a region on a map. Counties, states, lakes, oceans, and bays are examples of area data. Area data can be represented in a computer by describing their boundaries in a manner similar to the description of line data. (Note, however, that the boundaries may not always be clearly defined, as in the case of oceans.) Area data can also be represented by a quadtree [Samet 1982] or by tightly closed boundaries (TCB's) [Merrill 1973].

The three types of data can be generalized to one another, depending on the scale of the map. For example, a city, regarded as point data on a national map, may have to be treated as an area on a county map. The problem of map generalization is, however, not addressed in this paper.

<u>General Principles</u>
According to Imhof [Imhof 1975] there are six general principles that should be followed in annotating a map. They are:

1. Names should be easily readable and easily locatable.
2. A name and the object to which it belongs should be easily recognizable.
3. Covering, overlapping, and concealment should be avoided.
4. Names should assist directly in revealing spatial situation, territorial extent, connections, importance, and differentiation of objects.
5. Type arrangement should reflect the classification and hierarchy of objects on the map.
6. Names should not be evenly dispersed nor be densely clustered.

The automatic name placement program developed here was designed to adhere to these principles.

ANNOTATION RULES

In this section, the rules used by the automatic map annotation system are described. It should be pointed out that the rules given do not form a closed set; special rules may apply to special-purpose maps, and additional rules can be expected to be added as experience is gained with the system. For this reason it was regarded as important that the system be "rule-based" in its structure and facilitate the addition, modification, or deletion of annotation rules. There is, of course, no single set of annotation rules that is accepted by all cartographers. There are national and regional preferences, some agencies or organizations have particular rules of their own, and finally there are individual style preferences. An automatic annotation system must have the flexibility to accept specific rule sets, corresponding to particular agency standards, particular application requirements, or individual style preferences.

Basic Rules

The rules that apply to the labels of all types of map features are as follows:

(1) A name should not overlap another name or a point feature. If a name does overlap a line feature (or a boundary of an area feature), the line, not the name, should be interrupted.

(2) Names should not be evenly dispersed nor be densely clustered.

Area Feature Rules

The rules that apply to the name of area features are as follows:

(1) The label for an area feature should span the entire area and conform to the general shape of the feature, leaving about one and one-half letter spaces at both ends (see Fig. 1). However, if there is no significant difference between this placement and horizontal placement of an area

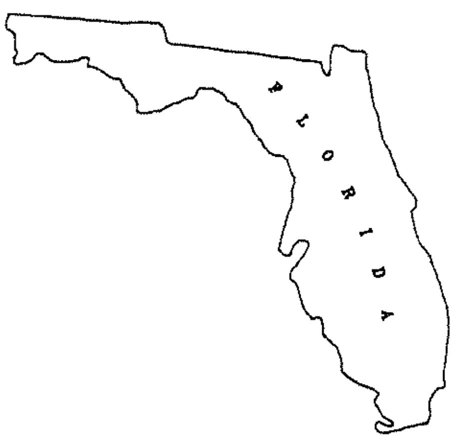


Figure 1. Name placement for an area feature.

name, then preference should be given to horizontal placement.

(2) Non-horizontal-placed names should not be straight, but curved. The arcs should not be greater than 60 degrees.

(3) A name that reads away from the horizontal is preferred over a name that reads toward the horizontal.

<u>Line Feature Rules</u>
The rules that apply to the name of line features are as follows:

(1) The label for a line feature should conform to the curvature of the line.
(2) Complicated and extreme curvatures should be avoided.
(3) Line feature labels should not be spread out, but may be repeated at reasonable intervals along the line.
(4) For horizontal line features, the names should be placed above the line. For vertical line features, there are two cases: If the line feature lies in the left half of the map, the name should be placed on the left side of the line to read upward. Otherwise, it should be placed on the right side of the line to read downward.
(5) One should avoid placing a name near an endpoint of the line feature.

<u>Point Feature Rules</u>
The following govern the placement of point feature names:

(1) The label for a point feature should be horizontal (usually east-west) and parallel to one of the map boundaries.
(2) Point feature labels, like line feature labels, should not be spread out.
(3) A point feature label should be close to the point feature to which it refers; though, some specified minimum distance must be maintained.
(4) Since the English language has many more "ascenders" than "descenders", "titles" (name labels that are above the point feature) are preferred to "signatures" (name labels that are below the data item).
(5) Although the best possible position of a point label is open to debate, Imhof recommends placement somewhat above and to the right of the point [Imhof 1975].

THE ANNOTATION ALGORITHM

<u>Basic Philosophy</u>
The basic requirement of any annotation system is that a name must unmistakably refer to the feature it designates and must not overlap point data. In order to achieve this goal, the following guideline was adopted: a name with a smaller degree of freedom is put in place before a name with a larger degree of freedom. This means that area features must be annotated first, since area names must be spread from one end to another, following the general shape of the area. When one considers that relatively large letters of an area name must be aligned with the other letters in the name while avoiding overlapping point data, it is easy to see that the degree of freedom for placing area names is relatively small. This is particularly true when the map is dense.

Point features have the next smallest degree of freedom since their names must be placed near the point to which the name refers. Line features can be labelled last since they have the greatest degree of freedom. Their names can be placed almost anywhere along the line; although, it is best to avoid placing a name too near the endpoints.

Area Feature Annotation

An area name should span the area to which it refers. This makes it necessary first to find a shape description of the area. It was decided to use the so-called skeleton method for this purpose [Montanari 1969]. The skeleton of an area is obtained as follows: the boundary layers of the area are peeled off one at a time, until the "wavefront" meets another wavefront. When no more layers can be peeled off, the skelton of the area will have been determined (see Fig. 2).

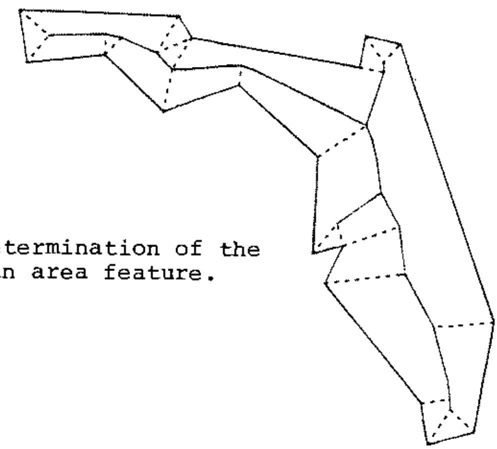


Figure 2. Determination of the skeleton of an area feature.

Before finding the skeleton of an area, however, a polygonal approximation must be made to smooth out small bays and peninsulas. This is required to prevent small shape variations from having an undue effect on the overall shape description.

Once the skeleton of an area has been determined, the skeleton is divided into sections by identifying the skeleton points whose associated values are less than a threshold that depends on the typesize for the name to be placed. These points are the "weak links" in the skeleton. All sections but the skeleton with the greatest associated area are deleted, and the skeleton path with the greatest associated area is determined. A straight, horizontal rectangle of fixed height and a curved non-horizontal "sausage" of fixed height are fitted to each non-disjoint subsection of this path. The heights of the rectangle and the "sausage" depend on the typesize of the name. The "sausage" or rectangle with the greatest area will be the area in which the name is placed.

If adjustments in the position of a name must be made because of conflicting data, they are made by either shifting the name perpendicular to the direction of the name, or altering the spacing between the letters. If a name is too large to fit into an area, it is placed outside the area, using the point-feature name placement algorithm.

Point Feature Annotation

To determine the optimum point-feature name placement, a graph of possible name positions is created. A node in the graph represents a point name. Two nodes are connected with a branch in the graph if their valid name placement areas overlap. To avoid comparing every node against every other node, nodes are sorted in order of increasing Y values. Then only the nodes that fall within a fixed Y range need to be compared against each other.

Once a graph has been constructed, it is divided into connected components by choosing any unprocessed node and processing all nodes connected to it, by traversing the connected component in a breath-first manner. This is repeated until no unprocessed nodes remain in the graph.

Once all the connected components have been determined, each is processed separately, since a node in one connected component can not effect the name label of a node in another.

For each node in a connected component, a list of free-space blocks is constructed by checking the positions of area names and neighboring point features. This is done by checking the grid cells that contain at least a part of the node's swept area. If an area name or a point feature falls within the existing free-space block, that free space block is split by removing the area overlapped by the area name or the point feature. If the resulting free-space block is too small to contain the point name, that free-space block is removed from the list.

Using the free-space list and the possible-positions list, a state-space search is carried out to place the point feature names. The initial state is the state in which no name has been placed for any point feature. The goal state is the state in which the names have been placed for all point features. The search algorithm used is a heuristic graph-searching algorithm similar to the well-known A* algorithm [Nilssen 1971]. However, it differs from the A* algorithm in that the nodes are ordered such that the node with the smallest degree of freedom is checked first, regardless of the physical position of the nodes. Since a larger name in a densely-clustered area will have almost no free-space, its degree of freedom will be small and such names will tend to be placed first.

When it becomes impossible to place a name, the algorithm backtracks, removing the names already placed to place them at different positions. Backtracking is helped by means of update records, which record the changes in the free-space blocks, the degrees of freedom for each node, and other internal information at the time the name labels are placed. Since the nodes are already sorted in the order of degrees of freedom, the amount of backtracking is relatively small in most cases.

Line Feature Annotation

Since a line feature may have to be labelled in more than one place, it is divided into segments of fixed length. Each segment is then labelled independently by testing every possible position of the name label for overlaps and selecting the best non-overlapping position. Starting at a certain distance from the end of the segment, sections of the line long enough to contain the name are searched until all possible positions have been considered.

When checking for overlap among name label positions of a line feature and an existing name label or point data, it is only necessary to check the side of the line which should contain the name label. To determine which side should be checked, a straight line is fitted to the section of the line corresponding to the name label position. If the line is vertical (or close to it), the position of the line section relative to the entire map is checked. If the line section is in the left (right) half of the map, the name should be on the left (right) side of the line section.

If the line is not vertical, the name should be above the line section. It is necessary to repeat this procedure for every line section since the line feature can change directions and the side of the line to be checked can change accordingly.

Two factors must be considered in deciding whether a particular position is a good position. The first factor, distance from the center of the line segment, is relatively simple to determine. The second factor relates to the curvature of the line at the particular name position. This is determined by computing the absolute deviation of points in the line from a straight line fitted to the line section. This value must be normalized by the number of points that constitute the line section before it can be combined with the distance from the center of the line segment to determine the "desirability" of the name position.

Post-Processing Editor

Although it is intended that the annotation system be completely automatic, it is clear that some provision must be included for interactive editing of the result. The purpose of such editing is to improve the appearance of the map, to correct data errors, and to correct possible mistakes made by the system.

CURRENT STATUS AND DISCUSSION

Currently, the area and point annotation algorithms have been implemented on a PRIME-750 computer with a program written in RATFOR, a FORTRAN preprocessor. Results obtained with the area annotation algorithm are shown for the western part of the United States in Fig. 3. An example of point-feature name placement is shown in Fig. 4. A map labelled with the point and area annotation algorithms working together is shown in Fig. 5.

Although it will be some time before an annotation algorithm will be able to compete in quality against an experienced cartographer, some of the results to date are impressive. A few of the names in Fig. 3 are, however, clearly not in the best position. This is most notable for Arizona, for which the name runs in a direction opposite to those of the adjacent states. This is due to the fact that the skeleton considers only the local properties of the area features. A more sophisticated algorithm, which includes a set of rules involving interaction with the names of neighboring areas, has been outlined and will be implemented in the next version of the program.

ACKNOWLEDGEMENT

The research reported here was supported by the National Science Foundation, Computer Engineering Program, under Grant ECS 81-11619.

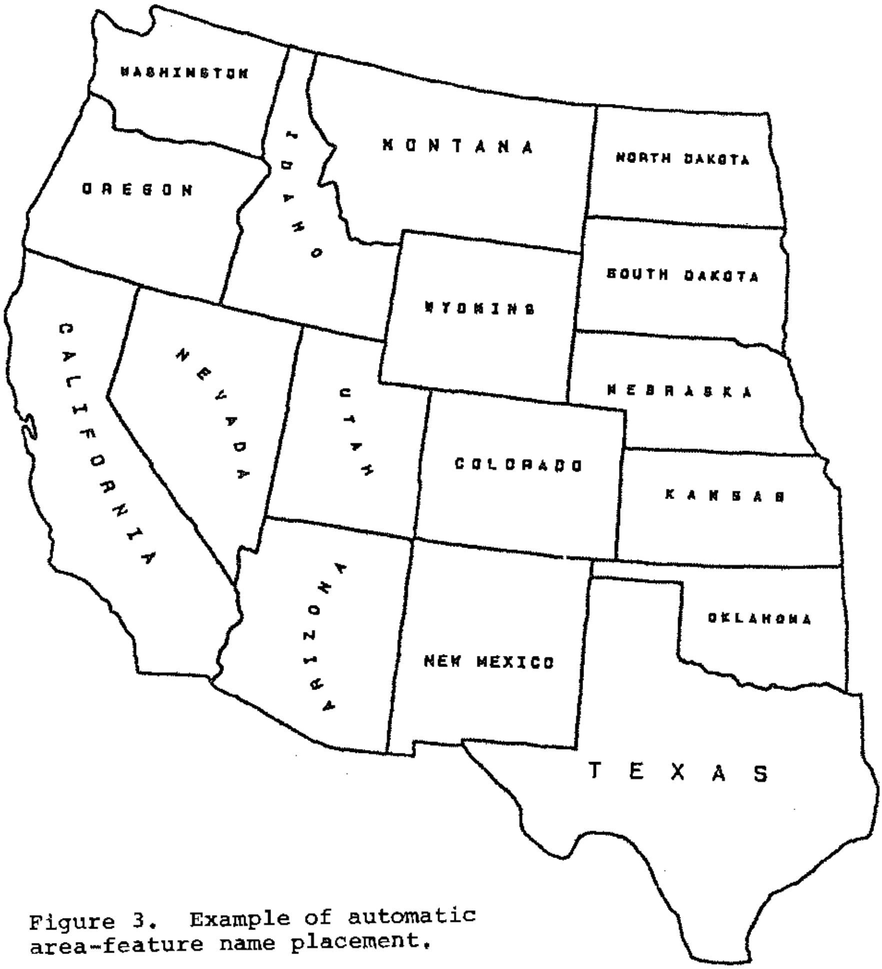


Figure 3. Example of automatic area-feature name placement.

REFERENCES

Freeman, H. 1974, Computer Processing of Line-Drawing Images, Computing Surveys, vol. 6, (1).

Hirsch, S.A. 1980, Algorithms for Automatic Name Placement of Point Data, MS thesis, Dept. of Geography, SUNY at Buffalo, Buffalo, NY.

Imhof, E. 1975, Positioning Names on Maps, The American Cartographer, vol. 2, (2).

Kelly, P.C. 1980, Automated Positioning of Feature Names on Maps, MS thesis, Dept. of Geography, SUNY at Buffalo, Buffalo, NY.

Merrill, R.D. 1973, Representation of Contours and Regions for Efficient Computer Search, CACM, vol. 16, (2).

Montanari, U. 1969, Continuous Skeletons from Digitized Images, JACM, vol. 16, (4).

Nilsson, N.J. 1971, Problem-Solving Methods in Artificial Intelligence, McGraw-Hill, New York.

Samet, H. 1982, Neighbor-Finding Technique for Images Represented by Quadtrees, CGIP, vol. 18, (1).

Figure 4. Example of point-feature name placement.

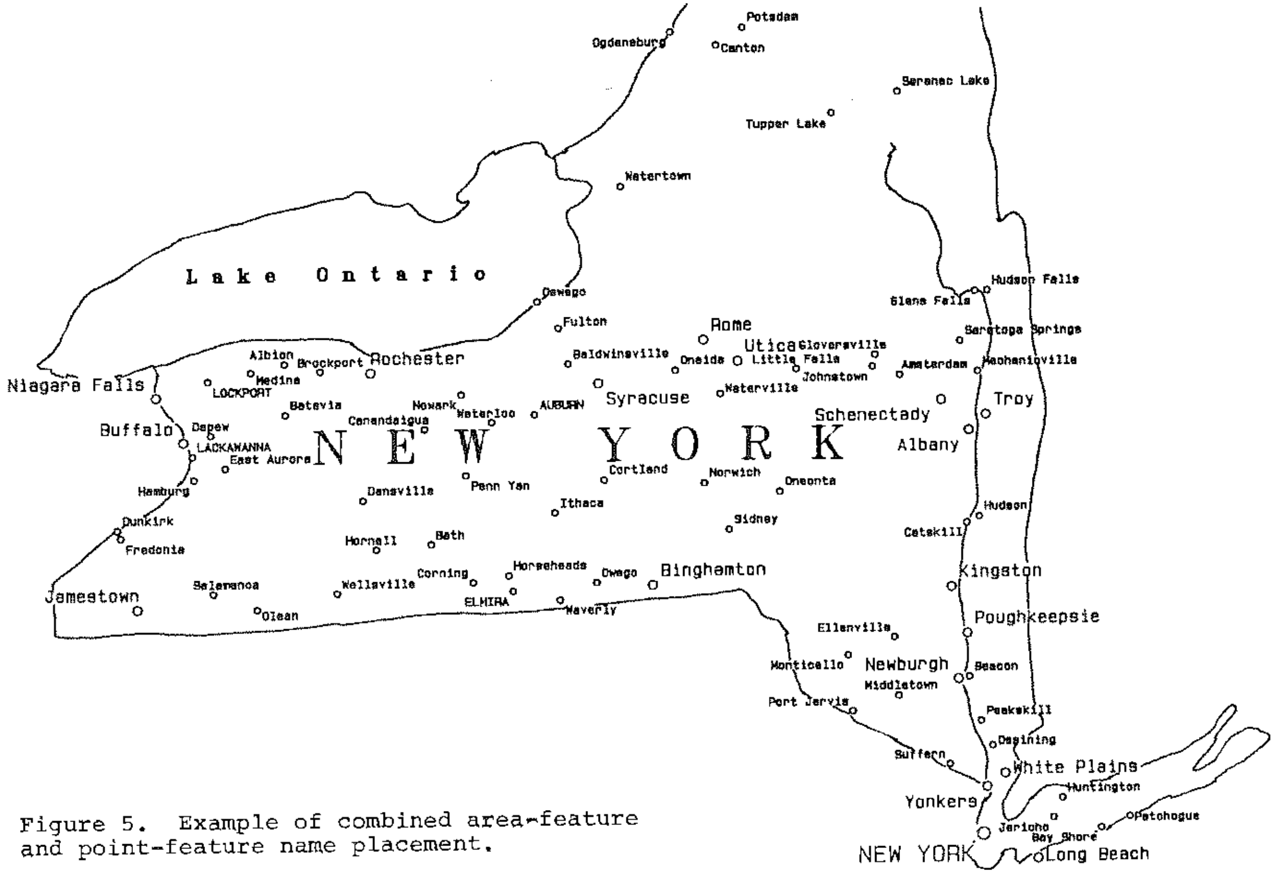


Figure 5. Example of combined area-feature and point-feature name placement.

THE MAKING OF THE FAR EASTERN ECONOMIC REVIEW ECONOMIC AND SOCIAL ATLAS OF THE PACIFIC BASIN

D.K. Forbes, C.C. Kissling, K. Mitchell, M.J. Taylor, N.J.Thrift and S.Wild

Department of Human Geography
Research School of Pacific Studies
Australian National University
GPO Box 4
Canberra, ACT 2601
Australia

ABSTRACT

This paper reviews the process and problems involved in producing a computer-based colour atlas of the Asia and Pacific region. Issues relating to graphics technology, data availability and production coordination are explored. While the requirement for annual or biennial updating made computer cartographic methods appealing, experience in meeting commercial deadlines lends only qualified support to the methods currently available in Australia.

INTRODUCTION

Most world atlases are eurocentric. In how many atlases is the Pacific Basin peripheral - relegated to the end of the volume and, to add insult to injury, split down the middle with New Zealand taking on a double identity. If this approach were ever to be justified it would have to be written in terms of the old colonial imperatives in which the countries of the Basin figured as the farthest reaches of several empires. But times have changed and now the countries of the Pacific Basin arguably form the major node of economic growth in the world today. It is therefore appropriate that an economic and social atlas should focus on this region. In November 1981 we conceived the idea of an economic and social atlas that would capture the major changes that have occurred in this region in the second development decade of the 1970s.

Recently a new generation of atlases has sprung up around the world. They have tried to cast off the straightjacket of the old academic type of atlas and communicate with a lay audience on immediate issues. Such atlases (for example Kidron and Segal 1981) have three major distinguishing characteristics. First, they are broad brush in the approach they adopt. They do not try to overwhelm the reader with a surfeit of detail. Second, they are concerned with immediate issues, with the result that the issues that are chosen to be mapped are usually economic, social and political indicators like inflation, indebtedness, aid or defence rather than mine by mine surveys of mineral deposits, field by field accounts of agriculture, plant by plant accounts of industry or container by container accounts of trade. Finally, in line with the need to communicate complex situations simply, maps have to be designed so they have an immediate impact. This means the use of bright primary colours, large lettering and careful attention to map design. To fill out local detail and provide a guide to further reading, it was decided that a 600 word text should accompany each map.

With these thoughts in mind it was obvious that we needed a large commercial publisher with an efficient marketing and distribution network. To bring an atlas on the region, out of the region we therefore entered into negotiations with the Hong Kong-based South China Morning Post Group to publish the volume under the auspices of the Far Eastern Economic Review. We were asked to provide sample maps and text. Four hand-drawn maps were produced and texts prepared which led to the signing of a contract in September 1982.

Between the inception of the project and the signing of the contract we had been exploring the possibilities of using automated cartography to produce the maps. There were possibilities of producing the atlas on the University of California, Santa Barbara graphics system, but we were fortunate to be able to enlist the aid of the Commonwealth Scientific and Industrial Research Organisation (CSIRO) in Canberra. CSIRO were willing to use the atlas project as a means of testing a colour-graphic system called Colorview which they were developing in collaboration with the proprietory Sydney-based firm TECHWAY. The publishers' response to this computer graphics option was enthusiastic because it held out the possibility of rapid updating of the atlas and republication on an annual or biennial basis. The very nature of the issues to be covered in the atlas enhanced the desirability of this option and made it a possible companion to the well-respected Far Eastern Economic Review, Asia Yearbook.

One of the problems of the new breed of economic and social atlas is that the kind of issues that are to be covered are also issues about which often very little is known. Apart from the more obvious economic and demographic indicators like GNP or life-expectancy data tends to be sparse and is rarely available from a single source. Therefore, one of the major focusses of the atlas endeavour has been the building up of a series of coherent consistent and, as near as possible, comprehensive data bases, most frequently from a variety of sources. Along with the preparation of the maps themselves, the assemblage and computerisation of the maps themselves proved to be the major task that faced us. This paper is arranged as follows: in section one we discuss the development of the map base and a suitable projection; in section two we consider the selection of data to be mapped and data collection; in section three we review the production process.

DEVELOPING THE MAP BASE PROJECTION

The project was instigated shortly after work had begun on computer cartographics in the Department of Human

Geography at the Australian National University, and was consequently a useful experiment in investigating various techniques.

At the time the project was started programs had already been developed to handle projection, rotation, windowing and generalisation of graphic data. These programs were being used in conjunction with World Data Bank 1 to plot maps of various parts of the world at different scales and on different projections. The maps were then used by either cartographers as base maps, or with further overlays of geographic data already on the computer, such as distributions of artefact locations.

For the Atlas project two maps were required:

1. A world map, showing coastal and national boundaries, which was to be overlayed later using conventional cartographic methods. The map was projected using Winkle's Triple projection centred on Latitude 130°E.

2. A regional map of South Asia, the western Pacific rim and the Pacific Islands, comprising coastlines and national boundaries. This map was to be completed using computer techniques to infil the different countries in various colours to show the quantity of the displayed attribute, such as the amount of money spent on defence in each country, for example. The map created used the azimuthal equidistant projection centered on 120°E, 6°N.

The infilling of the regional map was to be executed through the Colorview system. This system is based on defined polygons, which were either islands or areas of mainland, which were then collected together into countries. The collections of polygons could then be coloured according to specified attribute values to produce a map. The system was still under development by CSIRO with the atlas project being used as part of an evaluation process.

An important problem was the conversion of the projected map into the required polygon format for the Colorview system - zones, polygons, line segments and points. Using World Data Bank 1 this should have been straight forward as each line segment in the data bank is defined with a left and right polygon identifier. A program was written to access this information and, when used on test data for Europe worked well. It was then applied to the atlas region producing unbelievable results. On checking the original data various problems were discovered. Some of the left and right polygon identifiers were wrong and confusion arose over separate islands having identical polygon identifiers, for example the Australian mainland and Tasmania. At this stage a new program was started to try and avoid these problems, but owing to pressure of time it was decided to generate a large projected map of the atlas region on a pen plotter at ANU and to re-digitize it at CISRO with correct polygon definitions.

The digitizing undertaken at CSIRO was then in the format required for the Colorview system.

At ANU an interactive form-filling and editing program was adopted to receive the zone attribute data and convert it to the correct format. The files produced were then transferred by tape to CSIRO for the final generation of the maps. Again teething problems were encountered with the Colorview system.

SELECTION OF TOPICS AND ASSEMBLY OF DATA

Having decided that the theme of the Atlas was to be economic and social restructuring of Pacific Basin economies in the wake of a changing international division of labour, it was necessary to find adequate data sources to illustrate this process. Taking into account the sorts of data we knew to be available and the types of issues that needed to be dealt with, a list of 60 plates were developed (Table 1). Broadly the plates covered ten different aspects of socio-economic change. Four world maps sought to illustrate the place of Asian and Pacific economies in the world context, and these were followed by thirteen maps which highlighted significant aspects of economy and finance within the region. Following on from these were four maps on the structure of industrial production, five on energy and mineral resources, and seven maps on agricultural production. Then came three maps on the workforce, and ten on population issues, followed by four on human settlements and the environment, two on defence matters, and seven maps on transport and communication.

A wide variety of data sources was required. The growth of interest in the region as a whole, combined with an unexplainable tendency for several publishing houses to shift into the bibliographic guides market, meant there were several reasonably comprehensive new bibliographies of economic and financial data sources for the region (e.g. Blauvelt and Durlacher 1981). Broadly speaking, three main sources of statistical information were exploited. First, world sources of data published by organisations like the United Nations, the International Labour Office and the World Bank. Second, publications dealing with particular configurations of countries within the region. The Economic and Social Commission for Asia and the Pacific and the Far Eastern Economic Review cover all or most of the countries in the region, the South Pacific Commission is a valuable source for Pacific Island countries, the Central Intelligence Agency's National Foreign Assessment Center produces useful handbooks on socialist states and particular country sources are necessary for Taiwan and frequently China and the USSR. Third, special sources of information such as Amnesty International's reports on political prisoners and Boulding's Handbook of International Data on Women were consulted to fill in gaps on specific topics.

Despite the extensive published statistical sources available for the Atlas, major problems emerged in compiling

Plate no:

1 The countries of the region

Asia in the World Economy

2 Customer dependence in trade: exports
3 Customer dependence in trade: imports
4 Nationality of the world's largest enterprises
5 Nationality of the world's largest banks

Economy and Finance

6 Gross National Product per capita
7 Structure of GNP: agriculture
8 Structure of GNP: industry, construction and mining
9 Structure of GNP: services
10 Trade specialisation
11 Balance of trade
12 Foreign direct investment
13 External debt
14 Budget deficit
15 Inflation
16 Country risk
17 Aid receipts from all sources
18 Aid receipts from OECD countries

Industry

19 Production of vehicles and fabrics
20 Production of radio and television receivers
21 Shipbuilding and crude steel production
22 Export processing zones

Resources

23 Energy balance
24 Energy production
25 Energy reserves
26 Mineral production 1
27 Mineral production 2

Agriculture

28 Land use
29 Rice
30 Wheat
31 Maize
32 Barley and millet
33 Livestock
34 Products of the sea

Workforce

35 Workforce
36 Women in the workforce
37 Labour conflict

... cont'd

Plate no: (cont'd)

Population

38 Population and population change
39 Projected population in the year 2000
40 Birth rates and family planning
41 Mortality
42 Age structure
43 Primary education
44 Secondary education
45 Tertiary education
46 Muslims
47 Refugees

Human Settlements and the Environment

48 Level of urbanisation
49 Social indicators
50 Housing
51 Access to clean water

Defence

52 Military expenditure
53 Armed forces

Transport and Communication

54 Tourists and short-term visitors
55 International air traffic
56 Shipping
57 Daily newspapers
58 Television and radio
59 Telephone communication
60 Linguistic diversity

Table 1. Atlas topics

complete series for almost all maps. Sometimes inaccuracies were revealed by comparing figures from different sources, while at other times the poor data collection practices of the least developed countries in the region revealed itself in gaps and improbable estimates. These sorts of problems were compounded in the economic maps by the difficulty of measuring socialist economies using categories derived from capitalist economies. Working on the principle that the point of the Atlas was to illustrate and discuss a series of broad inter-related patterns, inadequate data was carefully evaluated and, if necessary discarded. Most of the remaining data was broken down into between two and five classes, on the assumption that this would minimize the distortions in the raw data.

Basically three forms of statistical data were collected. First, data sets were collected and stored in a computer file if they were to be used in the computer-graphics system; in other words, if they were to provide the colour-base to the maps (Table 2). Second, data sets for overlays were also stored in the computer file, but it was the hard-copy printout which was used for compiling the hand-drawn overlays. Third, supplementary data for the text accompanying the maps were generally stored by the author responsible for the particular map. Because it was selective data it was generally of a less variable quality than the comprehensive data needed for compiling the maps.

Map number: 435

Title: services & govt

Missing Value: -1.00 Units: % wkforce

Afghanistan	:	21.40
Australia	:	64.50
Bangladesh	:	6.50
Bhutan	:	3.70
Brunei	:	63.00
Burma	:	24.40
New Caledonia	:	49.00
Cambodia	:	9.50
China	:	9.00
Fiji	:	42.00
French Poly.	:	65.00
Hong Kong	:	49.70
India	:	10.00
Indonesia	:	27.90
Japan	:	55.60
Kiribati	:	70.00
Laos	:	10.10
Macau	:	46.10
Malaysia	:	37.60
Maldives	:	-1.00
Mongolia	:	59.90
N. Korea	:	10.10
Nepal	:	21.90
New Zealand	:	57.50
Pakistan	:	24.00
Papua - NG	:	50.00
Philippines	:	32.80
S. Korea	:	40.50
Singapore	:	64.20
Solomon Is.	:	42.00
Soviet Asia	:	36.00
Sri Lanka	:	37.80
Taiwan	:	39.20
Thailand	:	8.50
Tonga	:	31.00
Trust T.	:	72.00
Vanuatu	:	18.00
Vietnam	:	4.70
W. Samoa	:	31.00

Table 2. Data file for Plate 35, Workforce

PRODUCTION PROCESSES

The cartographic demand of the atlas was to produce approximately 90 coloured maps showing both one and two variable information (i.e. coloured background plus symbol overlay) within a time limit of six months. Because the proposed maps covered the same area with the same regional breakdowns the utilisation of the Colorview system that was currently being developed by CSIRO appeared to be ideal. It was anticipated that this method of production would produce a cost saving in both time and money. With problems inherent in developing systems the final time available, once the system was 99 per cent bug-free, was reduced to two months.

The initial limitation of the system was with the quality of linework that could be produced from the graphics equipment available. The quality of linework was not of a sufficient standard to produce a satisfactory final result. So, while it was practical to use the computer-graphics package for the background colour infils, conventional methods had to be used for linework and lettering.

The cartographic designer spent four weeks working with the CSIRO computer system - this period included familiarisation and bug and breakdown time. With the experience gained the same results could it is estimated be achieved within two weeks. The Colorview system in respect to the operator was found to be user friendly. With the system a choice could be made between twelve different quantitisation methods plus three colour selection methods. The majority of time was spent in colour selection, six sequences being selected using the blue/red/green method which were then filed away for instant recall when required.

The method employed by the cartographic desiqner, after logging on, was:

1. Outline of area called up onto colour screen ready for colour infilling.

2. The colour sequence to be used transferred from stored file.

3. The map data, previously transferred from ANU to CSIRO, called up.

4. The quantitisation method plus limits of each class division were selected to produce a map that was both informative and aesthetically pleasing.

5. This information was filed. From these files CSIRO then produced a colour positive at a 3"x3½" format which was sent to the publishers. These positives were enlarged and scanned to produce the required colour separations for printing.

The overlays were then drawn by conventional means. One basic overlay was produced for all one-variable maps. The two-variable maps required their own separate overlays.

A number of problems were encountered in using computer generated positives to create the final atlas maps and additional problems were encountered in connection with the combining of conventional and computer techniques.

1. Registration of colour bases and overlays proved problematic. Because the coloured positives were produced at 3"x3½", it was impractical to put registration marks on them. All overlays were drawn at 8"x9½" (final size) which caused the problems of registration to be taken out of the cartographic designer's hands and placed in those of the photo-lithographer and printer.

2. The amount of multiple photographic processing has been a major concern. Problems of quality control in photographic work especially when several parties at great distances apart are involved increases the likelihood of foreign matter being included and blemishes being generated. For the Atlas project computer generated film has first been processed in Canberra and then enlarged and scanned in Hong Kong, during which time minor and not so minor blemishes have appeared.

3. The faint image of raster lines on large areas of colour were an aesthetic problem. This problem was associated mainly with the sea area on the maps and was overcome by using conventional cartographic methods in the form of a computer generated mask of the sea.

4. Owing to the many processes that the colours have been put through, from the shades that were displayed on the VDU screen to the cartographic designer, to the processing of the positives, the enlarging and scanning of those positives and the final printing, we were not able to predict the exact final colour.

Thus, the cartographer's view of the method involved and of the processes that the selected colours must go through is that too much control of the quality of the final product has been taken out of the designer's hands and spread amongst too many groups. There appears to be a definite time-saving attraction to computer graphics through computer-generated overlay masks. This method would put quality control back into the cartographic designer's hands and would allow corrections and modifications to be made easily by hand prior to printing, obviating the need to regenerate an overlay or even the whole map if a colour positive is used.

CONCLUSIONS

In conclusion, computer graphics are ideal for producing the new generation of economic and social atlases of which the Far Eastern Economic Review Atlas is one example. Whilst it would be incorrect to say that there were no teething problems, once these problems are overcome graphics are an excellent and cheap way of updating economic and social information of contemporary importance as well as producing over time, an invaluable historical record.

We have come across some other problems to do with our institutional context. For example, a number of our colleagues seem to believe that this kind of atlas is "too commercial" or "not academic enough". We want to argue strongly and to the contrary that this kind of atlas often provides more information relevant to academics than the old kind of academic atlas and, furthermore, performs the invaluable task of keeping the public informed about economic and social change, one of the classical prerogatives of any liberal academic institution. We suspect that part of our colleagues' concern comes from precisely the fact that the type of atlas made possible by this kind of technology makes this information available to a wider audience, demystifies it and removes it from the ivory tower.

REFERENCES

Blauvelt, E and Durlacher, J. 1981, eds. Sources of Asian/Pacific Economic Information, Gower, Aldershot.

Kidron, M and Segal R. 1981, The State of the World Atlas, Simon and Schuster, New York.

TOWARDS AN ELECTRONIC ATLAS

Dr. Ewa Siekierska
Surveys and Mapping Branch
Department of Energy, Mines and Resources
615 Booth St., Ottawa, Canada, K1A 09E

ABSTRACT

This paper describes selected features of an experimental Graphic Work Station installed in the Geographical Services Directorate of the Surveys and Mapping Branch. The station was developed in order to test chosen cartographic and geographic methods required for the design and manipulation of electronic thematic maps. The cartographic functions included are various aspects of cartographic generalization, the selection of colours, and the structural design of visual planes. The geographical methods addressed are the interactive selection of class intervals, the simple functions for a simulation model and various techniques for data base queries.

INTRODUCTION

In 1981 the Geographical Services Directorate purchased hardware for an experimental Graphic Work Station (GWS). The main reason for this purchase was to develop an experimental system for thematic map processing. The system will be used for testing and evaluating the concepts and methods that, in the future, could be incorporated in a full-scale electronic atlas system. Such a system could support the Geographical Services Directorate in the design of conventional thematic maps, and could provide the capability for an extended form of cartographic presentation.

HARDWARE

The Graphic Work Station has the typical configuration of an interactive system for computer graphics. Its main processor is a micro-computer, the LSI 11/23 with 256Kb memory additionally supported by a large quantity of high-speed disc storage; a removable standard CDC disc with a 60Mb memory is used. The essential part of a graphic system is an interactive, intelligent, colour display. A device selected as sufficient for an experimental application is the raster graphic system, Lexidata System 3400, with a medium resolution display comparable to the standard North American television monitor. The System 3400 has a separate, programmable, high-speed memory controller and a microprocessor. It has 11 picture memories which can display selectively a map with 16 separate overlays. The Lexidata System chosen has a joystick for controlling interactive functions and a tablet for the input of summarized commands. The use of the graphic system reduces software development needs, since some operations, such as overlaying or colour manipulation, can be performed directly in the graphic system memory.

SOFTWARE

The contract to implement the Graphic Work Station was awarded to the Canada Systems Group (CSG), Advanced Technology Systems Division. In addition to writing the user requirements, the EMR development team worked in close cooperation with CSG personnel during the design, development, and implementation stages. Such cooperation was necessary for the development of the experimental system, so that the user specifications could be modified to take advantage of the specific characteristics of the hardware and the methods used could be insured to be consistent with the methodology of geography and cartography. The Graphic Work Station software is written in Pascal, except for a few short programs that are written in Assembler in order to optimize the processing of some operations.

DATA BASE DESIGN AND STRUCTURING OF DATA FOR PROCESSING

The Graphic Work Station data files can be subdivided into the following major file types: a) map feature coordinate files, in vector mode, b) thematic attribute files, associated with each feature, c) cartographic specification files, and d) temporary or permanent output files, in raster mode, which are used for the manipulation and the display of electronic maps on the graphic monitor. Figure 1 shows in more detail the relationships between the major GWS data files.

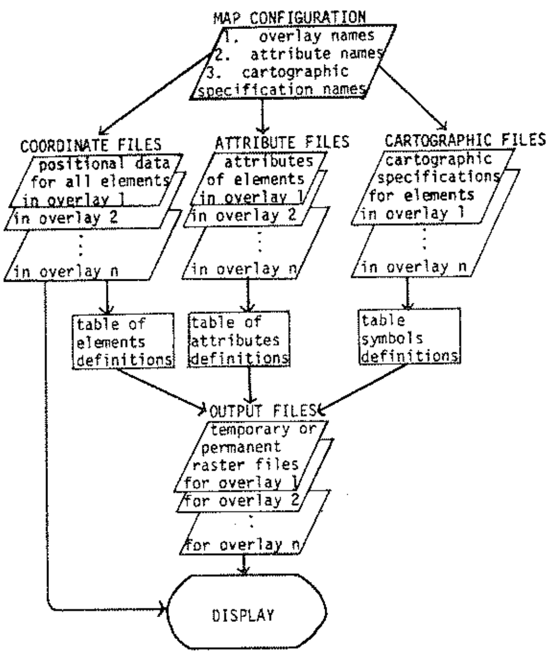


Fig. 1 Configuration of data files in Graphic Work Station

The names of files that have been chosen for a given map are specified in a general file called MAP CONFIGURATION. This file contains, firstly, the names of the features to be used for a given map (it calls COORDINATE files), secondly, the attributes of those features to be displayed (it calls ATTRIBUTE files) and, finally, it specifies how all of these will be represented (it calls CARTOGRAPHIC representation files). Each of these files is associated with definition tables, which store respectively: a) permissible element names, b) permissible types of attributes and c) permissible cartographic specifications. If the chosen MAP CONFIGURATION parameters are in agreement with the definition tables, then the vector coordinates, together with the chosen attributes and according to the selected cartographic specifications, are converted into a raster-mode, pixel array file, which is stored in a buffer (OUTPUT FILES) for further processing and displaying. Most of the operations in GWS are performed in raster format. The exceptions are line editing functions, which are performed directly on the vector files.

The logical division of the input data is consistent with the traditional division of maps into separate components for reproduction (cf. overlays in figure 2). In the Graphic Work Station, the homogeneous features of a given map component form one file and can be manipulated as one map overlay. For example, all small, populated places, represented with point symbols are stored in one file and can be displayed separately. Examples of combinations of overlays are given in the section on visual structuring of electronic maps. This organization of data proved to be very convenient for dynamic manipulation.

The information stored on the maps represents a vast amount of data. Therefore, to improve the efficiency of processing, both the vector and the raster data files have been subdivided into small map-subsections called "mapels". In the Lexidata System, the mapel size of 1024 x 1024 digitization units (one unit = 0.02 mm) was optimal for processing. Hence one mapel corresponds approximately to 2x2 cm on the input material, and 32 x 32 pixels on the display screen. Such a division of data files is simple and convenient for computers, and is the most natural for spatially referenced data (Kestner J. Oraas S. 1976).

INPUT DATA

The thematic material being used to build a preliminary Graphic Work Station data base comes from the 5th edition of The National Atlas of Canada (previous editions of The National Atlas were published in 1906, 1915, 1957, and 1974). The fifth edition breaks with the traditional practice of producing bound volumes of maps and will be issued instead as separate map sheets with individual covers (Groot 1979). The information contained in these maps will eventually be available in digital form. The subjects chosen for the experimental project are energy, agriculture, and population. The energy topics are oil fields, oil pipelines, and oil refineries; the population data are Indian and Inuit communities. The last group, namely agricultural lands and soil capability for agriculture, was not fully implemented at the time of writing this paper (June 1983). The subjects included in the data base are just a small selection of the topics dealt with in the National Atlas program (Falconer 1983). Aside from the attractivness of these topics to the Department of Energy, Mines and Resources, the chosen material represents the major types of symbols used for the presentation of thematic data. For example, oil refineries are point symbols applied to data of ordinal type. The oil pipelines represent quantitative flow line symbols at the interval-ratio level of measurement; the oil fields are examples of isolated areas and soil capability is of an interconnected-polygon type.

The thematic information is superimposed on the base map material. The Geographical Services Directorate is responsible for providing and maintaining up-to-date base maps for both its own use and use by other agencies. The base-map use is two-fold: it serves, firstly, as a general reference map, and secondly, as a reference framework for the thematic information. The base map components included in the GWS data base are: the graticule, rivers and lakes, provincial boundaries, roads, and selected populated places.

In the electronic atlas, a base map could serve the same purpose as in conventional cartography; but in a much more flexible fashion, since the user may select only those map components that are required for a given presentation. Moreover, this choice may be performed interactively, since visual evaluation is a very important factor in the design of a well balanced, visually pleasing, product.

CARTOGRAPHIC FUNCTIONS

Visual Structuring of Electronic Maps

Effective cartographic communication involves various methods for the visual structuring of map information (McCleary 1983). The Graphic Work Station provides a variety of functions for the designing and visual structuring of maps. One set of functions has been implemented for those who will use the system only for the manipulation of overlays already existing in GWS output files. These functions include the selection and display of individual map elements (or multiple combinations of such) or an enlargement of selected areas in a map inset. Another set of functions is implemented for users who would like to design their own maps. Those functions permit users to choose any set (or to create a subset) of input files and to display them according to their own specifications.

One important advantage of the raster display system for an electronic map is the existence of high-speed memory planes. Thus, besides displaying conventionally "complete" maps, it is also possible to use the information associated with each memory plane selectively. In the GWS, the smallest accessible display unit is a map cell. A map cell may consist of one or more map overlays, and can be presented using a particular set of colours. The present version of the GWS permits at most six overlays to be combined into one display cell, and one output map may consist of at most eight cells. The order in which the map overlays are listed in a given cell, and also the order of the cells themselves, define their visual priority on the monitor. Figures 2 and 3 illustrate two possible ways of organizing map overlays.

MAP STRUCTURE

MAP NAME: basemap

CELL#	# OF COLOURS	OVERLAY FILES
1	1	graticules
2	1	lakes, rivers
3	1	roads
4	1	boundary
5	2	city.s
6	2	city.l
7		
8		

Fig.2 Organization of overlays for base map manipulation

Figure 2 is an example of a flexible organization scheme where most overlays are individually accessible (the only exception is an overlay with hydrographic data, where lakes and rivers are handled concurrently).

MAP STRUCTURE

MAP NAME: oilproduction

CELL#	# OF COLOURS	OVERLAY FILES
1	1	graticules, lakes, rivers, city.s
2	2	roads, boundary, city.l
3	16	oil
.		
.		
.		
8		

Fig.3 Organization of overlays for thematic map manipulation

Figure 3 is an example of the scheme typically used for thematic maps. In this case, overlays with the base map components are grouped into two map cells, and will be displayed with only three colours. The majority of colours are reserved for the manipulation of the thematic attributes; whereas the base map will be used only as a reference.

Besides the manipulation of the individual memory planes, other features that are unique for electronic maps are the real time scroll and the instant enlargement functions. The scroll function helps to overcome the limitations of a relatively small sized display. It permits browsing over the areas adjacent to those actually displayed on the screen. The instant enlargement function is very advantageous for displaying maps. With the ever-increasing speed of computing, it may eventually be used for the creation of a scale-independent digital map. The GWS permits two-fold and three-fold scale enlargements with strict control over the magnification rules for cartographic elements. Thus only the symbols with real spatial dimensions, such as lakes or large cities, can be expanded during magnification; whereas the other elements, such as symbols for roads or graticules, remain unchanged. Towards this end, it should be emphasized that, for the proper presentation of a map, the use of a "mechanical zoom", which is unrelated to map scale, is not permitted in cartography, because in cartographic communication the presentation must closely relate to the accuracy of input data (see section on electronic map generalization).

Selection and Manipulation of Colours

Colour is one of the most important graphic variables in cartography. It increases map legibility significantly. The advantages of using an electronic display for thematic cartography began to be fully appreciated after the popularization of colour monitors. The black and white graphic displays have been used mainly for engineering plans and large scale maps.

The Lexidata System 3400 permits a large variety of colours. Each colour is created by specifying a combination of the additive colour components red, green, and blue (RGB) by giving percentages of each colour component to control their relative intensities. The colour value, as defined in cartography (Robinson Sale and Morrison 1978), is altered by adding or subtracting all three primary colours in equal proportions. The third dimension of colour, called in cartography intensity, is derived by manipulation of hue and value simultaneously. The System 3400 permits creation of 16.7 million colours. However, using a large number of colours on one map is not recommended, because the human visual system reacts efficiently to only a limited number of colours. In the GWS the number of colours needed for a given overlay is specified in the MAP STRUCTURE record, which is a part of CONFIGURATION FILE. The total number of colours possible for a given map is

$\sum_{i=1}^{8} n_i$, where $\sum_{i=1}^{8} n_i \leqslant 127$ and $\sum_{i=1}^{8} \lceil \log_2 n_i \rceil \leqslant 8$

where n_i is the number of colours in the i-th overlay and where $\lceil\ \rceil$ is the "integer ceiling" function.

The GWS provides three techniques for generating colours. One permits the individual specification of colours associated with a given map element, a set of elements, or their attributes. Another function permits the "automatic" creation of colour scale; and the last one permits an interactive manipulation of colour scales. To create an automatic colour scale the user sets parameters for the first and last colours in the scale, and the remaining colours are derived automatically in equal steps of colour value. The third method, which is implemented in the GWS, permits the interactive dynamic creation of a colour scale. In this function, each class is assigned a particular colour and the whole scale is interactively manipulated to obtain the desired size of class intervals (see section on interactive manipulation of class intervals) and the optimal combination of colours.

In order to relate to the conventional cartographic selection of colours, the graphic system should include the appropriate colour dictionaries. After testing the characteristics of a particular colour monitor, it is possible to calculate numerical colour transformations to standard colour systems such as the CIE or the Munsell scale. For the use of single colours, it is recommended that, using various perception tests, a set of "focal colours" be established, i.e., the best set of basic hues for a given monitor (Cowan 1983). The creation of the colour scales is best performed interactively by controlling all three colour dimensions simultaneously. From the colour scales created automatically on the GWS, the best results were obtained either using a grey-level scale or a combination of single hues with white or black (black with blue, white with green, and white with red). A linear progression of RGB components gives a better impression of equal steps than does a linear progression of colours obtained using printing screens, since the voltages controlling the colour brightness have similar characteristics to light perception curves.

As an additional variable, not available in the conventional map presentation, the graphic displays permit the use of a "blink" function, which is very advantageous for highlighting small areas. In the GWS this function is used extensively to identify overlapping regions when comparing features from separate overlays. Another similar function, which in future may be used for a rapid identification of symbols on an electronic map, is the use of motion. This function seems to be particularly relevant in enhancing the cartographic displacement technique used when locating symbols in "overcrowded" areas. A good example is the displacement of oil refineries in the city of Edmonton. Due to the size of the symbols used for refineries, they have been displaced around the city symbol. The use of motion, which will indicate the true location of the refineries, could significantly enhance the cartographic communication by electronic maps.

Some Aspects of Cartographic Generalization

Generalization belongs to one of the most frequently misused cartographic operations in the field of computer graphics or even in computer-assisted cartography. Very often this complex multi-dimensional process of map simplification (Robinson Sale and Morrison 1978) is reduced to the automation of one particular aspect, for example line smoothing. Generalization was one of the first problems addressed in computer-assisted cartography (Morse 1967) and still remains practically unsolved. At present, EMR conducts experiments to devise generalization methods, which will permit the automated derivation of topographical features from the 1:20 000 scale to update the 1:50 000 scale maps (Fraser 1981).

One aspect of generalization chosen for the first experiments in the GWS relates to the theory of visual perception thresholds (Ratajski 1978). This aspect seems particularly relevant to the graphic display systems, since most of them have an inbuilt "mechanical zoom" capability that permits the user to enlarge or reduce instantaneously the map image. This may result in serious distortions of the map's presentation since a "scale independent digital map" can not be displayed satisfactorily without generalization. Thus, one of the most important issues in electronic mapping is to design the methodology for enlarging or reducing display maps. Two approches are possible: the first is to use one data set and derive various maps using generalization procedures; the second is to create several data subsets, each associated with a band of scales each derived from a given data subset.

In the GWS the first approach was attempted, i.e., the automatic derivation of various scales from one data set. As mentioned in the section on input data, the GWS data base was obtained from maps in the 1:2M scales; however, for one overlay (Indian and Inuit Reserves), this scale was found too small to show the outlines of areas of all reserves. Therefore, maps at the scale 1:500 000 were also digitized, reduced to 1:2M, and merged with other files. Because of the disparities in the size of the reserve areas, they are represented with two different symbols. For the larger areas, the true extent of the area symbol is used, but for the smaller areas that are below "visibility threshold" (in GWS a threshold of 7 pixels was selected) a point symbol was more appropriate. The conversion of an areal symbol into a point symbol is a scale-dependent operation; thus, after magnification to the scale of 1:1M, fewer reserves are represented with point symbols than at the scale 1:2M.

A possible extension of this function could be the creation of more complex rules for the conversion of symbols. Such rules could involve not only the areal but also the thematic dimension of data; for example, one could convert only those areas having large populations, whereas the others could be left out. Such modifications are technically straightforward and are considered for implementation in the near future.

GEOGRAPHIC FUNCTIONS

Data Base Query

The cartographic functions discussed in the previous chapter refer to methods relevant mainly to visual presentation. If graphic presentation is not required or else the user wishes to obtain more information about map elements than actually presented on the map, the electronic maps permit a large spectrum of "geographical functions", some of which are not possible in conventional cartographic communication. In this category are functions that permit the user direct access to the data base.

In the GWS several types of data base queries are possible. If a user is interested in knowing all the attributes associated with a given map element, he may ask for the complete listing using either the computer terminal key-board or directly pointing to an element on the display using the graphic cursor. Another type of query permits the user to obtain a list of elements with specified criteria.

For example, one may select only those oil fields which produce 10 or 20 million cubic metres per year. Data base queries are unique to information systems, and enlarge map communication capabilities. Besides the generalized visual information, users have access to the data files, which are unaffected by transformations needed for graphical presentation.

Simulation Models

Another geographical function unique to the data base is the possibility of derivation of new information that did not exist in the input data but that has been calculated. This capability is particularily useful in data bases used for resource management and planning applications, since it permits the creation of a variety of simulation models.

Calculations in the GWS are limited to the basic arithmetic operations. These functions were used to calculate how long oil reserves will last based on the current production figures. The present version of software permits calculations using two variables at a time. However, the potentiality of the GWS to include more complex simulation models is virtually unlimited. One possible extension being considered is to use animation in the dynamic simulation.

Interactive Selection of Class Intervals

The last GWS function to be discussed in this paper is the selection of class intervals. This operation relates to both the geographic and the cartographic aspects of data bases. It can be used either for numerical or visual evaluation of quantitative characteristics of data. A large variety of methods exists for the selection of class intervals (Evans 1977). Most of these are available to computers in elementary statistical packages; for example, in SPSS (Statistical Package for Social Sciences). The choice of the most appropriate method depends strongly on the characteristics of the data to be mapped or analyzed.

The GWS supports several methods for selection of class intervals. It is possible to obtain class intervals of equal steps, or intervals based on mean and standard deviation values, or any user-specified classes. Additionally, the GWS provides a very flexible interactive function that can be used for the selection of class intervals or for the quantitative analysis of data. It is based on the real-time manipulation of the frequency distribution histogram, which is displayed in a map inset area along with three cursors. The histogram can be calculated from the full range of data values or from any subset that the user intends to analyze. The interactive manipulation is performed by changing the positions of those cursors that indicate the current position of class boundaries. Each modification of position is indicated by a change of colour of all elements whose values correspond to the histogram subdivision of data values. Additionally, the numerical values of the current positions of these class-boundary cursors are displayed in the map legend. The interactive manipulation of the histogram functions provides the means for a very flexible analysis of data. Besides the capability of displaying frequency histograms, it provides the means for visual analysis of spatial distributions, which is a very important factor in the proper selection of class intervals.

SUMMARY AND CONCLUSIONS

The functions discussed in the paper are just a selection of the capabilities of the GWS and give only an indication of the full potential of graphic systems for the design and manipulation of an electronic atlas. But these functions illustrate well the flexibility that the user can expect in the visual presentation and manipulation of map information in digital cartographic data bases. One important difference, when compared with a conventional cartographic presentation, is the capability of obtaining user-demanded or even user-created visual presentations, which will satisfy more user-specific demands. An electronic atlas will readily permit unlimited combinations of topics included in its data base and also an almost unlimited variety of ways to display them. However, the latter should be governed by rules for the most efficient cartographic communication. Aside from so flexible a visual presentation, the electronic atlas permits direct access to its data base. Further, it facilites a variety of video-numerical analyses for evaluation of the spatial, thematic, and temporal dimensions of data. In an electronic atlas, such analyses could be performed directly on input data and not on output media, unlike the case for most of the measurements performed in cartography. This will ensure a higher accuracy of results. The calculations could include simulation models for mapping future distributions, an uncommon practice in conventional cartography.

The flexibility of display and the analysis of the electronic map are not possible without an application-oriented organization of the data base. The conceptual division of the GWS data base into three types of files (positional data, geographic attributes, and cartographic specifications) proved to be very appropriate for the creation of the well-balanced cartographic system. Existing map-processing systems have been devoted mainly towards processing positional data (e.g., automated drafting systems) or towards processing the geographical attributes (e.g., spatial analysis systems). The cartographic aspect was usually reduced to the use of display methods standard in the computer graphics field. Thus more effort is needed towards developing a cartographic system suitable for the creation and utilization of electronic maps. An important aspect of such a system would be the structuring of data according to the presentation mode, one example being the GWS organization scheme. Further, a cartographic system will require standardized sets of symbology, needed for efficient visual communication and for a methodologically correct presentation, i.e., the use of symbols and display methods consistent with the nature of the data and with the type and the scale of the map. Another important aspect to be included in the cartographic system are rules for automatic scale change, i.e., rules for the presentation of data consistent with the map scale and for the derivation of other scales. The unique nature of the digital map is that its generalization process is reversible, for example it is possible to regain the details apparently eliminated in a visual presentation during scale reduction.

It is the hope of the author that this paper might be a contribution toward a better understanding of the nature of an electronic atlas, an indication of the potential of graphic systems for thematic cartography, and the expression of needs for a new methodology in assuring effective cartographic communication through an electronic map.

ACKNOWLEDGEMENTS

The author wishes to acknowledge the cooperation and contributions of the project manager (Roger M.Defoe) and the numerous other members of EMR and CSG personnel who contributed towards the success of the GWS project, and assisted in the preparation of this paper.

BIBLIOGRAPHY

Cowan W. 1983, National Research Council, Division of Physics, Ottawa, Ont., K1A 0R6.

Evans I. 1977, The Selection of Class Intervals, in Transactions of Institute of British Geographers, vol.2, pp.98-124.

Falconer G. 1983, National Geographical Mapping Program, internal EMR publication, 8p.

Fraser C.S. 1981, Report on the Derivation and Revision of Maps in the 1:50 000 NTS Series from Provincial Topographic Mapping, Final Report, EMR Contract Nr. 1451798, 21p.

Groot R. 1979, Canada's National Atlas Program in the Computer Era, Cartographica, Monograph No.23, pp.40-52.

Kestner J. Oraas S. 1976, GOMAD - an Interactive System for Hydrographic Charts, Proceedings of the 16-th annual symposium on "Canadian Hydrographic Charts", pp.63-72.

McCleary G. 1983, An Effective Graphic "Vocabulary", IEEE Computer Graphics and Applications, Vol.3, No.2, pp.46-53.

Morse S. 1967, Generalized computer techniques for the solution of contour-map problem, Ph.D thesis, N.York University, 256p.

Ratajski L. 1973, Discourse on Cartographic Generalization, Polish Cartographical Review, Vol.5, No.2, pp.49-55, No.3, pp.103-110.

Robinson A. Sale R. and Morrison Y. 1978, Elements of Cartography, Willey, 448p.

APPLICATION OF A MODEL OF DYNAMIC CARTOGRAPHY TO THE STUDY OF THE EVOLUTION OF POPULATION DENSITY IN SPAIN FROM 1900 TO 1981.

J.A. Cebrián de Miguel
Department of Human Geography
Complutense University of Madrid.
(Spain)

ABSTRACT

It is often convenient to have a plastic image of the temporal evolution of a spatial feature. The model discussed in this paper builds such an image. Using a pen plotter, the program draws a series of pictures, defining the situation at every instant. These pictures are block diagrams, the vertical dimension of which is a function of the value of the thematic characteristic, that is, for example, population density in this case. The only information available corresponds to the census years 1900, 1910, 1920,... 1981. This has made it necessary for us to establish a procedure of temporal interpolation which yields intermediate pictures. We have chosen a linear interpolation procedure. Later on, we used a video camera to film all of the pictures which had been drawn automatically by the plotter. The movie that we obtained is 80 seconds long, corresponding to a temporal scale of 1 second/year. Although the system that we have available is not very precise, it is sharp enough to allow the movie to show how the Spanish population tends to concentrate around a few industrial areas -Madrid, Barcelona, the Basque Country and Valencia- following a progressive rate which reaches its peak during the 1950's and 1960's.

INTRODUCTION

The first mention of the convenience of using dynamic cartography procedures dates back to the early 1960's (Thrower, 1959 and 1961).
Initially, techniques were proposed for the preparation, by hand, of the graphs which were later to make up the photographic frames of a movie sequence.
Shortly afterwards, towards the middle of the sixties, Cornwell and Robinson (1966) discuss the possibilities of using computer graphic representation techniques to create the movie sequences which illustrate the processes of space-time changes.
From this time until the mid-1970's, a great many projects were carried out in this field, covering very diverse topics:erosion processes, changes in population distribution, urban growth (Tobler, 1970), atmospheric pollution, climatological changes, seismic processes, the distribution of traffic accidents (Moellering, 1973, and 1973 bis), etc.
In most cases, the strategy followed is to automatically reproduce each photographic frame, either on a cathode ray screen which is filmed under control, or directly, by means

of a microfilm-plotter connected to a computer.
More recently, work has begun on the development of dynamic cartography systems in real time (Moellering, 1978), which makes it possible to control the filming of a scene while it is being produced. This is an important step forward in that it allows the movement to be obtained interactively, making it more expressive, without having to decide previously all of the control parameters for the recording and thus avoiding certain inherent risks.
The procedure presented in this paper was developed within the framework of my Ph.D., which was finished recently, and does not take into account the new possibilities of dynamic cartography techniques in real time, automatically reproducing on paper the photographic frames which are to be filmed later. The reason for this is that in order to establish a dynamic cartography system in real time, it is essential to have certain graphic peripherals to which we did not have access at the time. It is also necessary to define a set of programs to deal with the highly complex graphic data which, due to its present unmanageability, we did not consider logical to write.

GRAPHIC MOVEMENT AND HISTORICAL SEQUENCES

In order to create the graphic representation of the evolution over time of a specific phenomenon, it has always been necessary to characterize both orthogonal dimensions of the plane in the following way: the horizontal component makes up the scale of historical evolution and the vertical component, that of the magnitude of the event being considered.
In this way, the group of points $P=\{P_1, P_2, \ldots, P_n\}$, such that $P_i=(t_i, v_i)$, t_i being equal to the temporal moment and v_i being the magnitude of the phenomenon at instant t_i, or the line which joins these points, give a suitable representation, in discrete and continuous form, respectively, of the process of change being studied.
This system of representation has some very clear limitations, since, according to this description, it can be used only for assumptions of temporal evolution of one specific characteristic which affects a single individual or case.
If, on the other hand, we attempt to represent simultaneously the temporal evolution of a certain trait which affects various indivuals, it is necessary to resort to, either some type of conventional sign -using different broken lines (fig. 1)- or releasing a third dimension on the drawing plane, along which the various cases are aligned (fig.2).
The location of cases in the representations we have dealt with is clearly non-spatial, since we did not take into account their actual situation in significant points on the earth's surface. If we now attempt to obtain a representation of the time-space variation of a specific characteristic, we can confirm that there is no satisfactory procedure to visualize this event on a single graph.
In fact, if we represent the value of the characteristic by points on a spatial grid (figs. 3 and 4), the three coherent linear dimensions that we are able to distinguish on the sketch plane are inmediately saturated.
Therefore, in order to obtain a graphic representation of

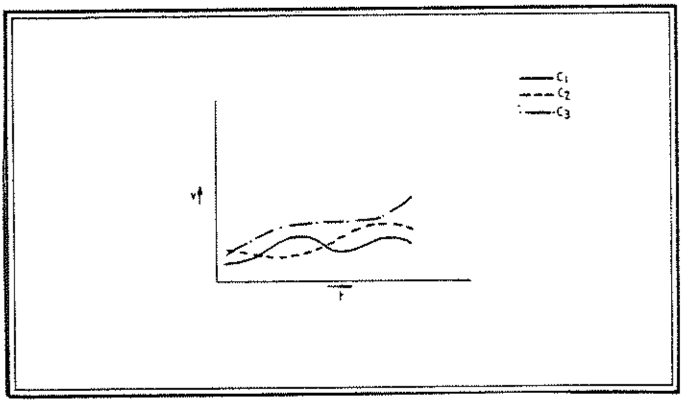


fig.1.

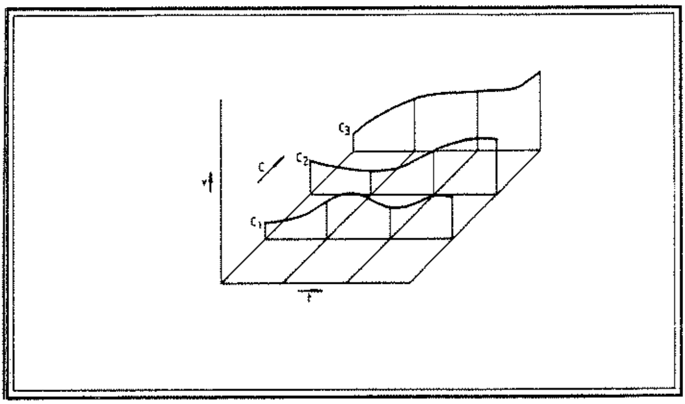


fig.2.

the space-time variation of a characteristic, it is necessary to resort to a series of drawings, each of which expresses the situation produced at a particular moment in time(fig.5).

For a long time, this was the usual procedure of presenting historical sequences of spatial characteristics, since the collection of thematic maps makes it possible to study the evolution over time of the spatial distribution of a specific event.

Dynamic cartography studies, on the other hand, assume a closer approximation to the visualization of the temporal evolution of spatial variables.

The basis for this type of treatment is very simple. The reason behind its complexity revolves around purely technical difficulties.

If, while maintaining the reference axes invariable, we film a collection of minimally dissimilar graphs, each one of which represents a moment in time of the distribution of a characteristic in the spatial sample, and if we present

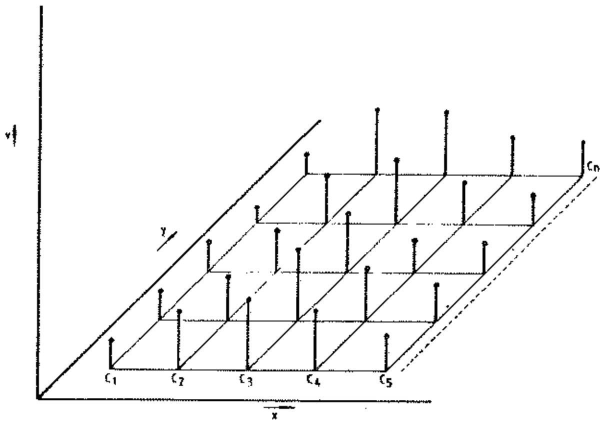


fig.3.

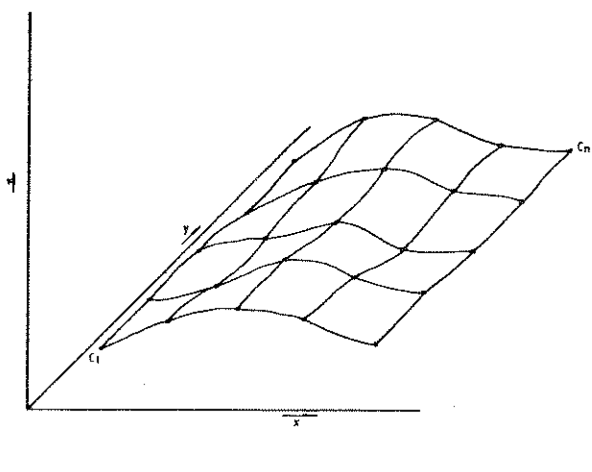


fig.4.

them for viewing at a rate of 24 images per second, the result will be the continous movement of a volume if, as in the present case, the type of representation selected is one of visualized block diagrams in isometric perspective. The successive configurations of this volume manifest, with a qualitatively different expressive force, the basic outline of a particular process of space-time evolution.

The first difficulty which must be overcome in order to ob-

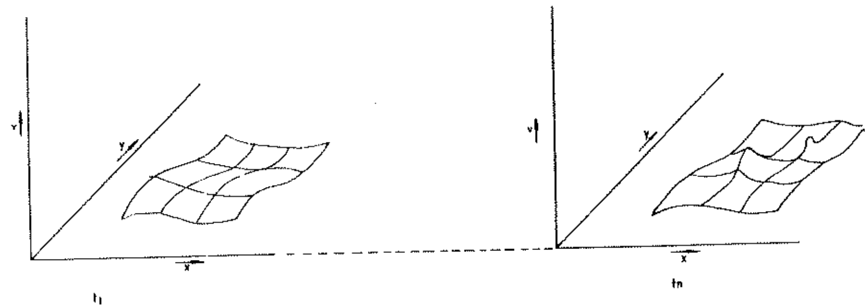


fig.5.

tain a moving sequence is the establishment of a large collection of images to be filmed. If automatic design procedures are not available, this pitfall is almost insurmountable.

The solution to the problem of the invariability during filming of the basic axes of reference is also practically impossible to reach if automatic control mechanisms are inavailable.

For the paper which we are presentig here, while we were able to solve the first problem, due to a lack of the appropiate machines, this was not the case for the second problem.

DEFINITION OF THE MOST SUITABLE MODEL OF THEMATIC REPRESENTAION

There are various possibilities available for the graphic representation of the values of a characteristic over a set of spatial individuals: choropleth maps, isopleth maps, conventional weighted sign maps, block diagrams in perspective, etc.

If a decision is made to present the variation in time of the distribution of such a characteristic by means of a collection of maps which correspond to a configuration at given moments, it is equally acceptable to use any one of the graphic representation procedures mentioned.

On the other hand, if we attempt to obtain a large collection of snapshots to later establish a moving sequence, not all of the procedures for representation are equally suitable.

The use of choropleth and isopleth maps is clearly inadvisable, since both types of representation, in order to be interpreted properly, require a considerable volume of alphanumerical information (keys on choropleth maps and labels for the level of the curves on isopleth maps). This

information is difficult to assimilate in a moving sequence.
Comparing the contributions made by the other two common procedures, we believe that the construction and representation in a perspective of thematic volumes is much more appropiate than the use of weighted conventional signs. The reason for this is that, while in the first procedure the variation of the characteristic is treated linearly, in the second case, this variation is associated to changes in superficial magnitude of the conventional symbols; thus, very small alterations of a linear magnitude are much more perceptible than those of a superficial magnitude.
These ideas have led us to use the procedure of representation by means of block diagrams in perspective. In this way, the moving sequence constitutes the visualization of the distortions of a volume, in accordance with the space-time variation of the phenomenon in question. In our opinion, this is undoubtedly the best solution.

DESCRIPTION OF THE 'FILM' PROCEDURE

Automatization of the Generation of Photographic Frames

In order to get a visual synthesis of the information in a matrix of values as

centroids	spatial coordinates		values of the characteristic V on the centroids in the T instants			
	X	Y	T_1	T_2	T_3 ...	T_N
C1	X1	Y1	V_{C1T1}	V_{C1T2}	V_{C1T3} ...	V_{C1TN}
C2	X2	Y2	V_{C2T1}	V_{C2T2}	V_{C2T3} ...	V_{C2TN}
C3	X3	Y3	V_{C3T1}	V_{C3T2}	V_{C3T3} ...	V_{C3TN}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
CM	XM	YM	V_{CMT1}	V_{CMT2}	V_{CMT3} ...	V_{CMTN}

we have defined a procedure of automatic generation of N block diagrams corresponding to the spatial distribution, at every single instant, of the characteristic V on each one of the M centroids, as defined by their cartesian coordinates.
As can be observed (fig.6), this procedure joins two basic subprograms: BLOCKD and INTEMP.
The first of these, using a pen plotter, builds and represents the thematic volume correspondig to the values of a characteristic -described in file V_{CTi}- over a set of centroids -defined by their cartesian coordinates in the CENTROIDS file- in a geographical domain -defined by cartesian coordinates in the OUTLINE file-.
Starting from the lists $/V_{CT1},V_{CT2}/$, $/V_{CT2},V_{CT3}/$,... $/V_{CTN-1},V_{CTN}/$, the second of these subprograms derives a number of list, V_{CTi}, equal to (N-1).NiM, NiM being a user-defined parameter. For each list, BLOCKD is called on to produce the graphic representation. All of the lists V_{CTi} are derived from each $/V_{CTN-1},V_{CTN}/$ by means of temporal linear interpolation and correspond to homogenous time

increments $\Delta T=(TN-TN-1)/NiM$.

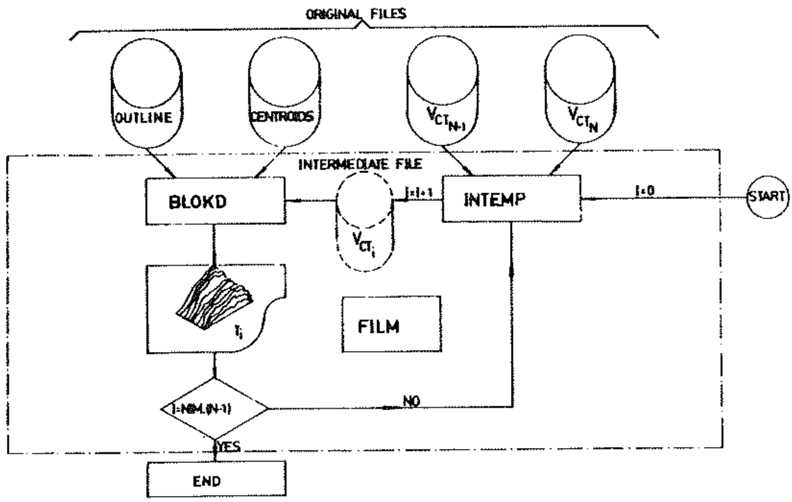


fig.6.

Filming of the Results
Due to the unavailability of a controlled filming system, the resulting drawings were recorded on a video camera. In spite of the fact that we used a reference panel and took great pains to keep these references from varying during the filming, this proved impossible to do given the means available. It is also possible to see a difference in the tonality of the lines , as the ink in the plotter pen was being used up during the process.
In this particular application which we developed, the resulting drawings were recorded at 24 frames/drawn and 12 frames/drawn.

EVOLUTION OF THE POPULATION DENSITIES
IN SPAIN FROM 1900 TO 1981

In order to check the contributions of this type of treatment in the analysis of space-time processes, we have carried out a moving sequence which depicts the evolution of the population densities in Spain between 1900 and 1981.
Beginning with the information on provincial population densities -compiled in the census publications- and with the digitization of the contour of peninsular Spain as well as the placement of the provincial capitals -taken as centroids in the representation- the FILM program reproduced eighty-one drawings, each one of which depicts the distribution, real or interpolated, of the population density for each year of the period. The dates corresponding to each graph were also registered, in order to obtain a temporal scale for interpretation of the moving sequence.
Two films were made of the plotter drawings, one of them at

a rate of 24 frames/drawn -time scale, 1 second/1 year- and the other at 12 frames/drawn -time scale, 0.5 second/1 year-. The following is a reproduction of one of the eighty-one drawings made by BLOCKD, which served as a basis for the filming (fig.7).

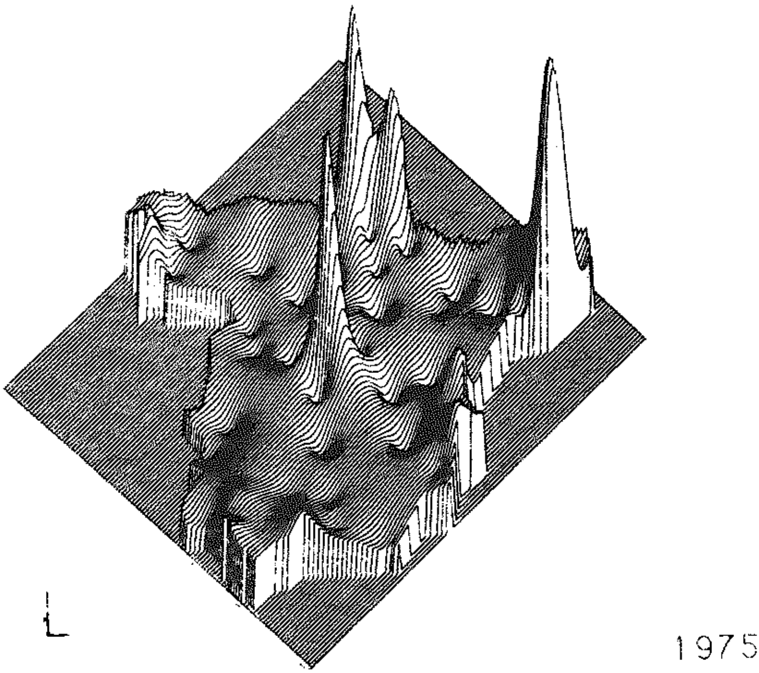


fig.7.

In spite of the abovementioned imperfections, it is easy to detect the progessive concentration of the Spanish population around a small number of industrial areas. This process reaches maximum intensity during the 1950's and 1960's.

ACKNOWLEDGMENTS

The autor wishes to thank Miguel García Ferrández for its assistance in the definition of the computer algorithms used, also José Ramón Colomines Boix in sequence filming, Beatriz Jiménez Blasco in the original data recording, and also Helmi Waits who translated this paper.

REFERENCES

Cornwell, B. & Robinson, A. 1966, Possibilities for Com-

puter Animated Films in Cartography: Cartographic Journal, 3, pp 79-82.
Moellering, H. 1973, The Computer Animated Film: A Dynamic Cartography: Proceedings of Association for Computing Machinery, pp 64-69.
Moellering, H. 1973bis, The potential uses of a Computer Animated Film in the Analysis of Geographical Patterns of Traffic Crashes: Accident Analysis and Prevention, 8, pp 215-227.
Moellering, H. 1978, An approach to the Real Time Cartographic Animation of three dimensional objects: Proceedings of the Ninth International Conference in Cartography, Maryland.
Thrower, N.J.W. 1959, Animated Cartography: Professional Geographer, 11, pp 9-12.
Thrower, N.J.W. 1961, Animated Cartography in the United States: International Yearbook of Cartography, pp 20-29.
Tobler, W. 1970, A computer movie simulating urban growth in Detroit Region: Economic Geography, 46, pp 32-38.

LE TRAÇAGE AUTOMATISÉ EN MODE NÉGATIF: LA COUCHE À TRACER ET LA COUCHE PELLICULABLE

Denis Genest
Société de Cartographie du Québec
1650, rue Louis-Jetté
Québec, QC Canada
G1S 2W3

ABSTRACT

The paper will first set a comparison between the negative and positive plotting modes and thereafter summarize the pros and cons of each of them. The selection criteria, for plotting equipment able to produce negative material, will be described in function of the particularities of the drafting material and tools that are commonly used for such plotting - link with traditional tools. The paper will then carry on in describing the current scribing techniques, outlining the difficulties that can be encountered and the way to avoid or correct them - techniques and tools. Finally, the mechanical preparation of peel-coat will be explained, highlighting the advantages of this method over the same work being done through the photo-mechanical process.

GÉNÉRALITÉS

Longtemps, les utilisateurs de cartes, particulièrement des cartes topographiques, se sont plaints de la surabondance d'information par rapport à leurs besoins particuliers. On aurait alors aimé pouvoir ne garder que les éléments vraiment pertinents mais, compte tenu des techniques traditionnelles de dessin, cela signifiait la préparation d'un nouveau calque, sélectif cette fois, et entraînait des délais considérables et une perte certaine de précision.

Maintenant, avec la constitution de fichiers informatisés qui contiennent la description, sous forme de codes et de coordonnées, des éléments de la carte, il est possible de faire le choix de l'information que l'on voudra voir apparaître sur «sa» carte. Par la suite, on fera tracer cette carte avec sa propre symbolisation et à l'échelle qui nous convient, tout cela, en relativement peu de temps.

En effet, les temps de traçage sont grandement diminués car, en premier lieu, les outils sont rapides dans leur fonctionnement: les divers moteurs déplacent la pointe traceuse à un bon train. De plus, comme les étapes d'interprétation et de généralisation, qui font appel à l'intelligence humaine, ont été accomplies lors de la construction des fichiers, il ne reste plus au traceur qu'à bêtement tracer, par déplacement en X et Y, en levant l'outil ou en le baissant, l'information qui est inscrite dans sa forme finale à l'intérieur du fichier informatique.

Cependant, et c'est là le sujet de cet exposé, on verra qu'il existe différents outils pour le traçage et qu'il en résulte des temps de traçage plus ou moins longs, en plus d'une qualité variable. Ainsi, nous tenterons de décrire les avantages du traçage en mode négatif sur le traçage en mode positif afin d'atteindre un but bien précis: la production de planches de dessin de très haute qualité, selon les normes usuelles en cartographie.

LES OUTILS DE TRAÇAGE AUTOMATIQUE

Il existe deux types d'images. D'abord, l'image où le sujet apparaît en couleur sur un fond blanc, transparent ou tout au moins de couleur plus pâle: c'est le positif, l'image à laquelle nous sommes le plus habitués (livres, journaux, oeuvres d'art,...) et celle que nous créons dès que nous prenons la plume.

A l'inverse, il y a le négatif, cette image qui nous surprend toujours lorsque nous recevons, en plus des photographies sur papier commandées au retour des vacances, ces bandes de plastique noirâtre où les cheveux sont plus pâles que les visages. Là, le sujet apparaît à l'inverse de la réalité, c'est-à-dire pâle sur l'image si normalement il est foncé, et vice versa.

Ainsi, actuellement, il est possible de préparer, par traçage automatique, des originaux positifs au moyen des outils suivants: stylos à bille, plumes feutrées, plumes mécaniques (régulières et pressurisée), traceurs électrostatiques, projecteurs optiques et faisceaux laser. Tous ces outils peuvent donner une image noire et, dans le cas des quatre premiers groupes, une image en couleur.

Par contre, on peut obtenir une image négative en employant soit des pointes de métal sur la couche à tracer, soit un couteau sur une couche pelliculable.

Comme on le verra, ces outils font appel à des tables traçantes et à des techniques plus ou moins sophistiquées. Il est quand même réconfortant de savoir que, quel que soit le produit recherché, il existe un outil et une technique pour l'atteindre.

TRAÇAGE EN MODE POSITIF VS TRAÇAGE EN MODE NEGATIF

Comme nous venons de le voir, le traçage en mode positif nous est le plus familier et s'effectue habituellement par dépôt d'une substance colorée, qu'il s'agisse d'encre, de peinture ou de tout autre liquide pigmenté. On peut aussi ajouter l'action de la lumière sur les émulsions photographiques qui donne, après développement, une image noire où la lumière a touché les sels d'argent.

A cause des techniques actuelles, des outils disponibles et des supports à dessin utilisés, le traçage en mode positif est plus rapide que le traçage en mode négatif. De plus, le dessin ainsi produit est directement exploitable: il ne fait pas partie d'un long processus de production et l'on peut s'en servir aussitôt que le traçage est terminé (...ou que le film a été développé, dans le cas des traçages optiques). Cependant, on ne pourrait envisager la production d'un même document à plus de cinq à dix exemplaires.

Si toutefois l'on veut intégrer dans le processus cartographique des positifs comme documents de base, on doit faire face à certains inconvénients. Ainsi, lorsque l'on doit tirer des copies photographiques d'un positif, cela exige par la suite une quantité imposante de retouches à faire sur le négatif obtenu, principalement à cause des poussières qui auront laissé leurs traces.

Les inconvénients du traçage à l'encre sont que, malgré que l'on puisse exploiter le traçage en couleur, les traits obtenus sont peu opaques et sont ainsi difficiles à copier. De plus, il est ardu de tracer des traits très fins et comme en cartographie, on fait la classification des éléments selon la largeur des traits qui les représentent, des traits de 0,08 mm, 0,10 mm, 0,13 mm, 0,15 mm et 0,20 mm se côtoient sur tous les documents et il est important que la largeur des traits soit la bonne. Il est impossible d'atteindre cette qualité de trait par un traçage à l'encre, d'abord parce que l'on ne fabrique pas de plumes selon cette variété de largeurs de traits et ensuite, à cause des facteurs d'humidité de l'air et d'encrassement de la plume, il est pratiquement impossible de conserver la même largeur de trait en traçant avec une même plume sur deux documents différents, voire sur le même document. De plus, pour les non-initiés au dessin, il devient particulièrement frustrant d'avoir à surveiller régulièrement le débit de l'encre qui circule dans la plume: ainsi, si une plume est au repos pendant quelques instants ou si l'on emploie une encre peu fluide, la plume se bouche et cela crée des problèmes, surtout si l'on ne s'en aperçoit pas immédiatement.

On peut par contre pallier au problème des largeurs de traits très fins tout en restant dans le mode positif: il s'agit d'effectuer le traçage avec des instruments optiques, où la résolution des traits est très grande et variée. Cependant, ces outils sont parmi les plus chers et le traçage doit s'effectuer sur un matériel enduit d'une émulsion photographique, qui est donc cher à l'achat et qui nécessite un traitement chimique après que le traçage a été effectué. Il faut donc avoir accès à une chambre noire, d'abord pour effectuer le traçage, et ensuite, pour faire le développement, ce qui alourdit le mécanisme de production.

Comme autre difficulté, il faut souligner le fait qu'en cartographie, l'information représentée est habituellement fort dense: il y a donc avantage à effectuer manuellement les corrections minimes, se gardant toujours la possibilité de retracer à nouveau la carte si la quantité de corrections à faire est trop grande. Or, sur un positif, il est difficile d'effacer des traits sans attaquer la base de polyester que l'on emploie habituellement. Lorsque la base a été atteinte, il devient très difficile de redessiner à cet endroit, l'encre cherchant à fuir sur toute l'étendue de la zone grattée, ce qui donne un pâté. En ce qui concerne les émulsions photographiques, on pourrait sûrement mettre au point une émulsion de type «WASH-OFF» pour le traçage numérique mais on resterait toujours aux prises avec le problème de retracer, obligatoirement à l'encre, puisqu'une fois que le film a été développé on ne peut le réexposer. En ce qui concerne le traçage optique, il n'y a guère d'autres choix que de retracer optiquement, au complet...

Ainsi, les difficultés du dessin en positif étaient bien connues et ce, bien avant l'avènement des tables traçantes automatiques. Au début du siècle, par hasard, on a découvert qu'après avoir gratté une émulsion photographique selon ce qui ressemblerait à un trait, une copie de ce film faisait apparaître le trait gratté comme étant tout aussi valable que ce qui avait été dessiné originellement. On a donc développé d'une part, une émulsion qui pourrait être facilement grattée et d'autre part, des outils qui pourraient servir à retirer l'émulsion selon une largeur déterminée: le tracé sur couche était né et avec lui, le traçage en mode négatif.

On peut donc obtenir des traits d'une très grande opacité puisque, où l'émulsion a été retirée, la lumière est transmise à près de 100% alors que le reste de la surface est opaque à, aussi, près de 100%. On a de plus un contrôle parfait sur la largeur des traits ainsi produits et l'on est assuré que cette largeur sera constante sur le même document. Enfin, comme la plus grande partie du dessin est opaque, il y aura très peu de retouches à effectuer sur des copies qui seraient tirées de ce négatif. On peut donc minimiser la dégradation de la qualité des traits tout au long du processus cartographique et fournir à l'imprimeur des planches-mères qu'il pourra reproduire sans difficulté sur ses plaques.

Comme on le verra, la couche à tracer est faite d'une base de vitre ou plus habituellement de polyester, qui est enduite d'une émulsion opaque. L'émulsion est disponible sous forme liquide, ce qui permet de l'appliquer sur des traits qui auraient déjà été tracés et ainsi de les boucher. Comme il s'agit du même type d'émulsion que celui dont est originellement enduite la base, on peut retracer dans la nouvelle émulsion, sans que cela paraisse sur le résultat final. Cela constitue donc un double avantage.

Il est de plus très facile d'ajouter, manuellement ou avec la table traçante, de nouveaux traits tout en étant assuré de la largeur de trait voulue.

Malgré les avantages du tracé sur couche, il faut cependant mentionner que l'on doit procéder plus lentement qu'en traçage en mode positif. Pour un même fichier de traçage, le tracé sur couche prend 150% du temps requis pour la préparation d'un positif.

LA COUCHE À TRACER ET LES OUTILS REQUIS

La couche à tracer est faite d'une base transparente, de vitre ou plus communément de polyester, dont l'épaisseur est suffisante pour assurer la stabilité dimensionnelle de l'ensemble du dessin. Par la suite, cette base est enduite d'une émulsion chimique qui adhère à la base mais dont l'épaisseur est beaucoup moindre que celle de la base. L'émulsion peut être de différentes couleurs, ce qui donne différents degrés d'opacité par rapport à la lumière incidente.

On peut ensuite retirer l'émulsion de la base au moyen de pointes de métal: au cours du grattage, l'émulsion se transforme en poussière qu'il est facile de nettoyer avec une brosse. Un trait transparent résulte de l'opération et il s'agit de s'assurer que le trait sera au bon endroit et aura la bonne largeur.

Ainsi, on doit porter une certaine attention à la fabrication des pointes à tracer. Présentement, on trouve deux types de pointes. Tout d'abord, les pointes rondes offrent une zone de contact circulaire: la largeur du trait résultant sera égale au diamètre du cercle de contact. Ensuite, il y a les pointes tournevis qui, par analogie à l'outil bien connu, prennent la forme d'une arête linéaire: la largeur de trait sera égale à la longueur de l'arête, à condition que l'outil soit toujours perpendiculaire à l'axe de la ligne à tracer.

Alors que les pointes tournevis ont constamment besoin d'être orientées, les pointes rondes donnent toujours un trait de même largeur quelle que soit leur orientation. Toutefois, parce qu'elles offrent une surface de contact peu tranchante, les pointes rondes ne sont exploitables que lorsqu'elles sont minces. Il devient alors impérieux d'utiliser un système de contrôle de l'orientation des pointes tournevis, ce que peu de tables traçantes offrent.

Il faut de plus que la surface de traçage sur laquelle on étendra la couche à tracer soit très rigide et uniforme. A propos de la rigidité, il faut que la pression, exercée par la pointe sur l'émulsion, soit constante sur toute la surface traçable sinon la pointe pourrait perdre contact avec l'émulsion ou encore attaquer la base. Pour les mêmes raisons, il est primordial que la surface de traçage soit parfaitement plane. Comme les pointes sont taillées à partir d'un cône, toute variation de contact entre l'émulsion et la pointe résulte en une variation de largeur du trait, ce que l'on veut éviter en optant pour le tracé sur couche. Il est désirable, en plus, que la table traçante soit munie d'un système de succion afin de garder un contact étroit entre la surface de traçage et la couche à tracer.

On remarquera aussi que chaque pointe à tracer donne un meilleur rendement à des vitesse et pression qui lui sont propres. Ainsi, compte tenu du type de couche à tracer que l'on emploie, il faut établir quels sont ces paramètres pour chaque pointe et surtout, avoir la possibilité de faire varier la vitesse et la pression exercées.

Il y a enfin avantage à pouvoir positionner l'outil traceur à un endroit désiré, avec une très grande précision: on emploie alors une loupe munie d'un réticule et, à l'aide de mouvements infinitésimaux (0,05 mm) dans chacun des axes, on se situe à l'endroit voulu. Ceci peut devenir important si l'on veut ajouter de l'information supplémentaire sur une planche déjà tracée ou encore si l'on veut séparer l'information sur différentes planches qui seront mises en repérage.

DIFFICULTÉS RENCONTRÉES ET SOLUTIONS

Comme toute autre technique, le tracé sur couche requiert un apprentissage. Il reste cependant certaines difficultés qui demandent plus que de la pratique et de la persévérance. Il faut alors l'aide des manufacturiers.

En premier lieu, il y a le problème des départs mal définis: ceci arrive lorsque la pointe à tracer vient bel et bien en contact avec l'émulsion, dès le point de départ de l'élément à tracer, mais parce que la pointe n'est assez tranchante, l'émulsion n'est pas retirée complètement ce qui donne l'impression que la pointe «flotte» avant de pénétrer dans l'émulsion. Ceci se produit particulièrement avec des pointes rondes qui ont un diamètre trop grand.

Pour contourner cette difficulté, on peut soit employer des pointes tournevis, qui sont plus tranchantes, soit encore programmer un double passage dans la portion de trait qui correspond au départ. Dans le second cas, ce double passage ralentit le traçage en plus de nécessiter un traitement supplémentaire des données.

Une autre difficulté naît du dépôt d'émulsion dans les jonctions de traits. Ceci est dû au fait que les pointes tournevis, parce qu'elles sont très tranchantes, attaquent la base lorsqu'il n'y a plus d'émulsion à retirer. Cela se produit particulièrement où des traits se chevauchent, alors que la pointe va «sauter» par-dessus le trait déjà existant, laissant une partie d'émulsion non grattée de l'autre côté du trait déjà tracé. Ce problème peut devenir particulièrement aigu lorsque l'on trace des quadrillages, où chaque ligne chevauche presque toutes les autres.

Une solution immédiate consiste à faire tracer d'abord les traits les plus larges, puis les traits les plus minces. On minimise ainsi les résidus.

Toutefois, on a développé un nouveau type de pointes tournevis, où l'on a arrondi l'arête afin qu'elle soit moins tranchante et attaque moins cette base. Il y a bien entendu risque d'avoir des départs moins nets, mais dans le cas particulier des quadrillages où toutes les lignes se chevauchent, il y a toujours une ligne plus large qui recoupe le départ des lignes plus minces. Ces glissements des pointes tournevis arrondies sont quand même minimes et ne peuvent être facilement détectés; ainsi, la grande partie du tracé sur couche est fait avec ces dernières pointes à cause de leur fiabilité et de la qualité de traits qu'elles fournissent.

PREPARATION DE COUCHES PELLICULABLES

Jusqu'à date, nous avons parlé du traçage d'éléments linéaires: lignes droites ou courbes, il y avait toujours un outil capable de tracer cet élément sur toute sa longueur, donnant la largeur de trait désirée en ne passant qu'une seule fois. En ce qui concerne le dessin d'éléments représentant une superficie, p. ex. une maison ou une forêt, il deviendrait fastidieux et dangereux pour la qualité du résultat de prévoir un passage répété d'un outil traceur à l'intérieur de cette même étendue.

On a ainsi développé l'idée des masques ou fenêtres, qui correspondent à des zones à remplir d'une trame ou d'une couleur pleine. L'évidage de ces régions ne se fait pas en retirant progressivement toute l'émulsion d'un négatif ou en remplissant d'encre un positif, mais plutôt en découpant le contour de cette région et en enlevant par la suite une pellicule de matériel opaque qui recouvrait cette région.

Cette planche s'appelle une couche pelliculable et ressemble beaucoup à une couche à tracer. Cependant, c'est l'émulsion qui est différente: on ne peut la gratter car elle ne peut se transformer en poussière; il faut plutôt la couper avant de pouvoir retirer la portion désirée.

Il y a deux techniques pour couper l'émulsion. La première fait appel à une émulsion photographique que l'on ajoute par-dessus l'émulsion de la couche pelliculable. On expose cette planche sensibilisée à un document de référence, p. ex. la planche des maisons, et après développement on brûle, sur la couche pelliculable, les traits qui ont été exposés au moyen d'une solution chimique. On se retrouve donc pratiquement avec une copie conforme du négatif de base sauf que le type d'émulsion a changé: il s'agit d'une émulsion que l'on peut peler.

Il faut toutefois remarquer que le document de base peut contenir plus de traits que ceux dont nous avons besoin pour délimiter les régions à dénuder: il faut donc boucher ces traits et s'assurer de ne pas peler des étendues qui ne doivent pas l'être. De plus, comme nous ne voulons pas que la trame ou la couleur destinées à la surface dénudée aillent déborder dans le trait qui limite la surface, il faut aussi boucher ces traits. Toutefois, à cause de la largeur des traits qui délimitent la zone à peler, l'émulsion pelliculable, correspondant à la zone à peler, est plus petite que la zone, d'où naît le danger de créer des espaces blancs s'il y fautes de repérage.

Compte tenu de ces inconvénients et aussi à cause de la nécessité d'avoir recours à nouveau à un laboratoire photographique, on a pensé à découper directement sur la couche pelliculable, au moyen d'une lame, les régions à peler. Manuellement, c'est une tâche fastidieuse, qui correspond à un nouveau calque; mais si c'est effectué par une table traçante, on a pu sélectionner les éléments nécessaires à la délimitation des régions et c'est simple de remplacer une pointe à tracer par un couteau et de faire faire ainsi le découpage désiré.

Les avantages de cette méthode mécanique sont grands: le trait laissé par le couteau est infinitésimalement mince et comme l'émulsion pelliculable est élastique, le trait se referme sur lui-même après le passage du couteau et il n'est donc pas nécessaire de boucher les traits. De plus, la pellicule qui correspond à la zone à peler couvre toute la zone ce qui est important. Enfin, l'on contourne les besoins d'un laboratoire photographique, ce qui élimine des coûts et accélère le processus.

CONCLUSION

Le traçage en mode négatif doit être envisagé si l'on a besoin d'un traçage de haute qualité, avec des traits très fins et qui s'intégrera dans un processus cartographique devant mener à l'impression de documents en tirage plus ou moins fort. Il ne faut pas en attendre le document qui pourra aller immédiatement dans les mains de l'utilisateur mais plutôt le document d'assise qui servira à la production complexe d'une carte.

Toujours à l'intérieur du processus cartographique, il est maintenant possible de préparer des masques au moyen de couches pelliculables découpées avec une lame. Les traceurs numériques donnent d'excellents résultats tout en accélérant le processus de préparation et de pelliculage.

Ainsi, le cartographe a à sa disposition des outils qui lui permettent de s'affranchir des tâches les plus pénibles pour se consacrer à sa vraie responsabilité: gérer et représenter les données qui décrivent une situation.

BIBLIOGRAPHIE

Keates, J.S. 1973, Cartographic design and production, Longman, London.

Kongsberg vapenfabrikk, Technical Literature.

Oxtoby, P, Cartographic Techniques, I.T.C. published lecture notes, Enschede.

Robinson, Sale, Morrison 1978, Elements of Cartography, 4e, John Wiley & Sons, New York.

COMPUTER TO MAP

AN

EXERCISE IN COMMUNICATION

Rosemary Ommer
Department of History

and

Clifford Wood
Department of Geography

Memorial University of Newfoundland
St. John's

ABSTRACT

Preparation of graphics involving computer programmers, map author, and cartographer, gives rise to a series of communication problems. This presentation begins with an explanation of 19th Century historical records as raw data sources and how they were computer processed. From there it moves on to demonstrate the unexpected enormity of the data-set and its complexity. Thereafter the presentation examines the methods that were investigated in aggregating the important information to isolate and graphically illustrate the break point between success and failure in the 19th Century Eastern Canadian wooden sailing ship industry.

REMOTE SENSING: SYSTEMS AND ACTIVITIES

The Analysis of Landsat Imagery Using an Expert System: Forestry Applications 493
M. Goldberg, M. Alvo and G. Karam

LANDSAT Digital Data for Updating Glaciological Information on Topographic and Glacier Inventory Maps 504
Philip Howarth and Simon Ommanney

Enhancements and Classifications of LANDSAT Data for the Ecological Resource Survey and Mapping of the Aishihik Lake Environmentally Significant Area (ESA), Yukon Territory, Canada 514
Dave Puittinen and A.B. Kesik

La cartographie thématique en télédétection 524
J.R. Bélanger

A High-Accuracy Airborne Digital Line Imager 536
R.A. Neville, S.M. Till, W.D. McColl and J.R. Gibson

Analytical Plotter for Facility Management Systems Data Acquisition 541
R.R. "Sky" Chamard

Underwater Mapping Techniques Using Remote Sensing for Salvaging Sunken Vessels 549
A.A. Navarro

Computer-Assisted Photo-Interpretation (Abstract) 559
Carl Reed

Automatic Cartography of Agriculture Zone by Means of Multi-Temporal Segmentation of Remote Sensing Images (Abstract) 560
Eric Cals, Robert Jeansoulin and J. Claude Darcos

Limitations of LANDSAT Imagery for Thematic Urban Mapping (Abstract) 561
Mohammad Id Ozone

THE ANALYSIS OF LANDSAT IMAGERY USING AN EXPERT SYSTEM: FORESTRY APPLICATIONS

M. Goldberg[1], M. Alvo[2], G. Karam[3]
[1]Dept. of Elec. Eng., [2]Dept. of Math., Univ. of Ottawa
[3]Dept. of Syst. and Comp. Eng., Carleton Univ.
Ottawa, Canada

ABSTRACT

A production rule-based expert system for the analysis of multi-temporal LANDSAT imagery is described. Specifically, the system has been trained to analyze a forested region in Newfoundland, Canada with data acquired over a six year period. The main properties of production rule-based expert systems are first outlined; the various components involved - primitive predicates, derived predicates, goals, the rule base, and the control structure are described. The system developed handles the data in a sequential manner, each step in the sequence corresponding to one image. After classification of the newly acquired image, rules based upon the present classification, the previous classification, the previous decision, and various measures of confidence (or reliability) are invoked by the data. A new decision results for each pixel. The production rules are supplied by the expert user and can be modified as the need arises. Results are shown to indicate the utility of the expert system.

INTRODUCTION

With the launch of the American satellite LANDSAT in 1972, digital four-dimensional scanner imagery has become available. LANDSAT gives repetitive and complete coverage of the earth on an 18 day cycle so that it is feasible to use data from different times to either improve the accuracy of the classification and/or to detect changes in the forest cover. LANDSAT data has been used successfully in agriculture and in forestry for classification and identification of ground cover. As well, it has been used in the forming and updating of forest resource maps for regions that would be difficult to survey by conventional methods. However, automated methods for using LANDSAT data to monitor change brought about by logging, fire, regeneration or insect defoliation have not been developed.

The quick detection and monitoring of change in forests can become a valuable tool for forest management. Such a tool can assist government and industry in establishing accurate forest inventories.

For example, the cost of maintaining complete, up-to-date forest inventory maps over large areas (eg. 5000 sq. km) is enormous. By using inexpensive LANDSAT imagery to monitor forest change, timber inventory maps could be kept current (with some sacrifice of detail). Furthermore, detailed forest mapping (by air and ground surveys) can be directed only to areas where change has been detected, thus permitting more frequent mapping for a given budget.

In this study, we propose to demonstrate the feasibility of using LANDSAT data to detect and monitor changes in the forest environment. An area of Newfoundland of approximately 50 km by 50 km, corresponding to 1000 by 1000 pixel, was chosen, and the time period to be studied is from 1975 to 1981.

The problem of change detection from satellite imagery involves analyzing classified data from two images. Intuitively speaking, a change will be said to have occured if a significant number of pixels in one image have changed classes in the second image. Any change detection technique will have to alert the user to those changes and enable him to locate them precisely. In addition, since in practice some changes will be detected in classified pixels even though none have actually taken place, it will be necessary to adopt a technique which accounts for the effect of noise in the system.

The change detection system proposed in this study is based on the concept of an expert system. The system is presented with the classified imagery, reliability values, as well as the results of a statistical detection algorithm. In turn, the system, aided by a set of rules,interprets this information and subsequently makes a decision regarding the present classification of each pixel. The set of rules is predetermined by the user and it can be chosen to reflect the permissible changes. In the following sections, we describe in more detail the expert system.

EXPERT SYSTEM FUNDAMENTALS

One recent trend in artificial intelligence is the development of the so-called "expert system" for solving specific problem (Winston, 1977; Barr et al., 1981). In these systems the computer program is provided with the knowledge of experts in some field. The program can then make use of this knowledge for solving problems. In this system, knowledge is represented as a series of "production rules" (Winston, 1977) defines production rules as follows:

1. A production is a rule consisting of a situation recognition part and an action part. Thus a production is a situation-action pair in which the so-called left side is a list of things to watch for and the so-called right side is a list of things to do.

2. When productions are used in deductive systems, the situations that trigger productions are specified combinations of facts.

The actions are restricted to being assertions of new facts deduced directly from the triggering combination. The productions may be called premise conclusion pairs rather than situation-action pairs.

These production rules can be conveniently represented by "if ... then ..." statements; for example, if (P1 and P2, or P3) then P4.

Here the Ps are called predicates and are simply facts or conclusions. The above rule is interpreted as follows: If P1 and P2 are both true or P3 is true, then assume that P4 is true. Note that P4 can now be used as a predicate in some other production rule. Three types of predicates can occur:

1. Primitives: these are the basic facts or observations which are supplied to the system.

2. Decisions or Goals: these are the terminal predicates.

3. Derived predicates: the predicates are obtained from the right hand side of production rules.

Knowledge is represented by a series of "if then" statements. Intelligence is represented by the control structure of the program which decides which production rules to invoke, and in what order.

In (Winston, 1977) the following advantages of a production-systems approach are cited:

1. Production systems enforce a homogeneous representation of knowledge.

2. Production systems allow incremental growth through the addition of individual productions.

3. Production systems allow unplanned but useful interactions which are not possible with control structures in which all procedure interactions are determined beforehand. A piece of knowledge can be applied whenever appropriate, not just whenever a programmer predicts it can be appropriate.

IMPLEMENTATION OF THE FORESTRY EXPERT SYSTEM

In this section the FORESTRY EXPERT SYSTEM is described. This system is designed to treat the observations in a sequential manner; that is, one image at a time. For each new LANDSAT image there are two phases; image classification, and decision updating by the expert system. The first phase involves using pattern recognition concepts to classify the pixels and assign some measure of confidence or reliability of the classification to each pixel (Alvo et al., 1981). The next phase involves the FORESTRY EXPERT SYSTEM itself. Using it's rule base upon the new classification and reliability, the previous classification and reliability, and the previous decision, a new decision is derived. The flow of the analysis is illustrated in Fig.1. The imagery processed by the FORESTRY EXPERT SYSTEM is stored in an image database. The image database is designed to allow the expert system to select imagery for analysis without operator intervention. Descriptions of the image database organization and the expert system components will now follow. The expert system components will be expressed in the notation of a production rule-based system.

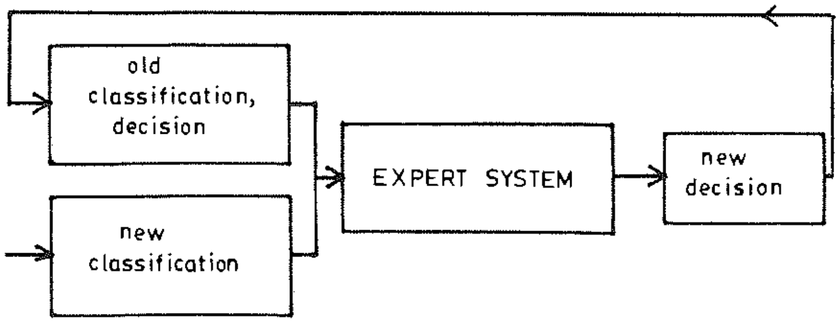


Figure 1. Flow of Analysis using the FORESTRY EXPERT SYSTEM

Image Database Organization

The fundamental purpose of any database is the storage of information with the intention of, at some time, retrieving and/or modifying the information. In this particular application the information is digital imagery; that is, any data represented by a line-pixel location, and an intensity value. In the expert system there is the further burden that the information retrieval must be performed (for the most part) automatically (not directed by an operator). This infers that the data description system used to access the different images must be unambiguous. When considering image data derived from or related to LANDSAT imagery, the selection of descriptors is vast. To determine those necessary, one must examine the exact requirements of the expert system.

The forestry expert analyzes images chronologically; furthermore, it must perform this using a number of different image types all derived from the same LANDSAT image, hence from the same image date. For example, the expert would compare a classification and reliability from one image date against the classification and reliability from the next chronological image. The two descriptors, data type and image date should be sufficient to enable the expert system to differentiate between images stored in the data base.

The image database consists of three components: (1) the image data, (2) the image (or channel) directory, and (3) the image descriptor blocks (IDB), one for each directory location.

The image directory denotes the presence of an image stored at a given channel number. In addition, if an image is being stored, the directory value corresponding to the image points to its IDB. The last six locations of the directory hold the "state" of the database; that is, which channels contain the latest decision, its reliability and the summer and winter classifications that were last used (and their associated reliabilities).

The IDB is a 150 byte descriptive record. It contains the data type of the image (a two digit code), the image date (if applicable), the image generation date (the date it was stored in the database), the image tape number (if applicable), and 80 bytes of descriptive text. The remaining 50 bytes are intended for future use.

Forestry Expert System

The FORESTRY EXPERT SYSTEM consists of 3 components: (1) the predicates, (2) the production rules, and (3) the control structure. Each of these are now described in some detail.

The predicates can exist in three forms: (1) primitive predicates (or simply primitives), (2) derived predicates, and (3) decisions (or conclusions). The primitive predicates are pixel or global predicates. The pixel primitives involve statements about individual pixels. Global primitives concern information common to a subset of pixels. The pixel primitives are: the previous classification C_t, the new classification C_{t+1}, and the respective classification reliabilities R_t and R_{t+1}. Since the decision-making structure is sequential, then the previous decision D_t, and its corresponding reliability DR_t are also considered pixel primitives.

Global primitives are based upon global statistics concerning the classification and reliability transistions from the previous classification to the next. The classification transition statistics are generated by calculating the number of pixels that are involved in a transition from a class C_{ti} to a class C_{t+1j}. Let this value be represented as C_{ij}.

Example: There are 100 pixels in class 1 of a 1975 summer image. In the 1976 summer image these pixels are redistributed as 75 pixels in class 1, 20 pixels in class 3, and 5 pixels in 8. Therefore the proportions indicated in the transistions for class "i" where i=1 are: (1,1)=0.75, (1,3)=0.20, (1,8)=0.05.

The reliability transition statistics are generated by counting the number of pixels in a transition C_{ij} such that $R_{t+1j} > R_{ti}$. Let this value be represented as R_{ij}.

Example: Suppose that the number of pixels in the (1,6) transition of two winter classifications was 75. Of these 75 pixels, 40 were more reliably classified in the newer classification than the older classification. The value in the reliability transition matrix at transition (1,6) would be 0.40.

The derived predicates are TRUE and FALSE logical values that are the result from the first type of production rules. The decision predicates are the "action" predicates from the second type of production rules. The execution of the expert system for one step of the sequence is concluded when these predicates are generated. Recall that one step of the sequence corresponds to the analysis of one LANDSAT image.

The production rules are executed in two phases; phase 1 using type 1 rules, and phase 2 using type 2 rules. The type 1 rules involve a statistical change detection inference and a reliability measure change inference. The purpose of these rules is to decide whether the changes in classification that occur are due to real changes or to misclassification. The expert must supply the values λ_{ij} and α_{ij} which indicate the degree of misclassification to be expected. These are represented mathematically as follows:

(1) STATISTICAL CHANGE DETECTION INFERENCE:

if $(C_{ij}/C_{ti} > \lambda_{ij})$ then C_{ij} is a statistically significant change, or, $P1(C_{ij})$ is TRUE - else - $P1(C_{ij})$ is FALSE

(2) RELIABILITY MEASURE CHANGE INFERENCE:

if $(R_{ij}/C_{ij} > \alpha_{ij})$ then R_{ij} is a reliable change, or, $P_2(R_{ij})$ is TRUE - else - $P2(R_{ij})$ is FALSE.

It should be noted that the comparison, $C_{ij}/C_{ti} > \lambda_{ij}$ is a simplistic representation of the actual calculation. In fact, there is a confidence interval associated with the comparison. These production rules can be represented by 2 tables, one table corresponding to each type of inference. The rows and columns of the tables specify the C_t and C_{t+1} labels, respectively. The table entries are the TRUE/FALSE logical values (or derived predicates). For example, if a group of pixels changed from class 1 in the previous image to class 5 of the new image, and the number of pixels that changed was statistically significant, then the statistical change detection table would contain the value 'TRUE' at location (1,5).

The application of the second type of production rules generates the decision predicates and these production rules assume the following general form:

(1) TO GENERATE THE NEXT DECISION:

if $(f_1\{D_{t_k}, C_{t_i}, C_{t\ ij}, P_1(C_{ij}), P_2(R_{ij})\})$ then D_{t+1_k}

(2) TO GENERATE THE NEXT DECISION RELIABILITY:

if $(f_2\{DR_{t_k}, R_{t_i}, R_{t\ ij}, P_1(C_{ij}), P_2(R_{ij})\})$ then DR_{t+1_k}

where: f_1 and f_2 are logical functions, and the general expressions are valid for all t. By grouping the rules according to the derived predicates P_1 and P_2, the type two production rule format can be simplified to the following: if $(P1(C_{ij})$ is TRUE) then if $(f\{D_t, C_t, C_{t+1}\})$ then D_{t+1}; $DR_{t+1} := R_{t+1}$; else if $(P2(R_{ij})$ is TRUE) then if ($f\{D_t, C_t, C_{t+1}\}$) then D_{t+1}; DR_{t+1}: = R_{t+1}; else D_{t+1}:= D_t;

if ($g\{DR_t, R_t, R_{t+1}\}$) then DR_{t+1};

It should be noted that, the production rules in the simplified form can be represented as functions of three variables.

An example will help clarify how the two phases of production rules function. Let us suppose the following situation: pixel "X" is classified (or observed) as softwood (class label 2) in a 1975 summer image, there i=2 for C_{ti}; the previous decision on pixel X is that of softwood (decision 10), or k=10 for D_{tk}; the next classification of pixel X labels it as a clear-cut area (class 1), that is j=1 for $C_{t+1,j}$; statistical change detection inference states that C_{21} is a statistically significant change, or Pl (C_{21}) is TRUE. A new decision can be arrived at in the following way: "Given that the previous decision was softwood, and a statistically significant change has occured, specifically softwood to clear-cut, it can be concluded that the area represented by pixel X has undergone logging (decision 5), or k=5 for $D_{t+1,k}$. Mathematically, the simplified production rule used in this example is:
if ($C_{t,2}$ and $C_{t+1,1}$ and $D_{t,10}$) then $D_{t+1,5}$.

The simplified type two production rules can be stored as a set of tables such as the one illustrated in Fig. 2. Each table represents the production rules associated with some $D_{t,k}$. The rows and columns of the tables are the C_t and C_{t+1} labels, respectively. The table entries are the new decision labels or $D_{t+1,k}$ values.

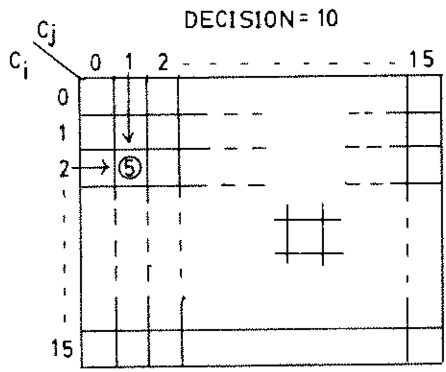


Figure 2. Decision Table Used for Simplified Production Rules

The use of table storage for the second phase production rules has several advantages. First, it permits easy entry or modification of the rules. Furthermore, the rules can be data driven; that is, they can be invoked when the proper combination of input data is available. In fact, the appropriate production rule for a set of data is found in one search.

In general, all of the second phase production rules cannot and should not be specified beforehand. Therefore, when a production rule that does not provide an "action" is invoked, then an "anomalous situation" is said to have occurred. Pixels that are affected shall be referred to as "anomalies". The anomalies are flagged by the system for an expert's interpretation. Once the expert arrives at some conclusion, the "action" part of the production rule can be specified. This involves entering a decision label in the appropriate table location. This is the procedure for "training" the expert system. The rule can now be re-invoked and the decision process completed. The only anomalies considered are those that originate when production rules are invoked in the presence of statistically significant change. Anomalies resulting when production rules are invoked by reliability measure changes, are generally too small in quantity to justify being monitored by the system.

The control structure for each interation of the expert system is now apparent. Execution begins with the primitive predicates, and proceeds through two phases of production rules until the new decision predicates are found. The structure is feedforward in design; that is, production rules never lead to "lower levels" of predicates. Once a decision is reached, the system state is advanced forward by one increment: $D_t \leftarrow D_{t+1}$, $DR_t \leftarrow DR_{t+1}$, $C_t \leftarrow C_{t+1}$, $R_t \leftarrow R_{t+1}$. When the next classification arrives, (chronologically greater than C_t), the system starts execution again from the primitive predicates.

In the next section, tests demonstrating the usefulness of the FORESTRY EXPERT SYSTEM are presented.

FORESTRY EXPERT SYSTEM TEST PROCEDURE AND RESULTS

The utility of the FORESTRY EXPERT SYSTEM is demonstrated in three areas:

1. Its ability to track profound forest changes such as logging and forest fire.

2. Its ability to detect changes only discernable using multi-temporal analysis; examples include forest regeneration and defoliation due to insect infestation.

3. Its ability to correct classification error caused by a variety of reasons and generally improve classification reliability.

The test scene employed was a 100 sq. km. area located in central Newfoundland, Canada. Over a period of six years the above mentioned man-made and naturally occurring environmental changes took place in this vicinity. The source imagery consisted of five geometrically corrected, classified LANDSAT scenes. The chronological sequence of the imagery is as follows: Dec. 16, 1975 (winter), Sept. 25, 1978 (summer), Feb. 16, 1979 (winter), June 4, 1979 (summer), and July 15, 1981 (summer).

The test results include the following:

1. Extensive logging took place progressively over the six year period. This is well indicated by the decisions resulting from the first four images (Fig. 3(a)(b)(c)(d)). The fifth image (Fig. 3(e)) recorded a forest fire that occluded much of the logged area.

2. Forest regeneration became apparent in the fifth image (Fig. 3(e)) as sections of the oldest logged areas began to regrow. This was inferred by a change from clear-cut to Not Sufficiently Restocked (NSRX) with a previous decision of logging.

3. Extensive, severe, spruce budworm insect damage was detected in the 1979 winter image (Fig. 4(a)). This was inferred from changes in winter imagery that indicated previously classified softwood areas appearing more like hardwoods. The insect damage spread after 1979 but was less concentrated and more sporadic (Fig. 4(b)(c)).

4. Classification correction is demonstrated in Fig. 5(a)(b) where weather phenomena caused a water area to be misclassified as a burn area. The knowledge of the expert system rejected this and retained the original decision.

5. Several areas of pure softwood stands were consistently classified the same in all five images. This increased confidence in classification was reflected by increased decision reliabilities for the pixels concerned.

All of the above phenomena have been verified by ground truth information from the Newfoundland Forest Research Centre (NFRC), in St. John's, Newfoundland, Canada. The ground truth data consisted of forest inventory maps interpreted from low level aircraft infrared photographs, and actual ground sampling.

CONCLUSIONS AND FURTHER RESEARCH

In this paper, a production rule-based expert system for the multi-temporal analysis of LANDSAT imagery has been specified and test results presented. Furthermore, the value of using a production rule-based approach has been demonstrated.

It has been shown that the expert's knowledge can be conveniently organized by production rules. In addition, the expert systems knowledge can be increased or corrected as new information becomes available.

Multi-temporal analysis of a 100 sq. km. area was performed using the FORESTRY EXPERT SYSTEM. The test results indicated that the system could track both highly discernable and very subtle forest changes. The results were verified correct by ground truth information. The use of the expert system to correct classification errors and generally improve the reliability of image classifications was also demonstrated.

Future work includes testing the FORESTRY EXPERT SYSTEM on a 2500 sq. km. area and developing a software structure for implementing complex production rules.

ACKNOWLEDGEMENTS

We would like to thank M. Haddad for generating the image classifications used in this research. In addition, we thank Dr. P. Kourtz of ENVIRONMENT CANADA for his advice and support.

REFERENCES

Alvo, M., Goldberg, M., 1981, "A Measure of Reliability for Classification of Earth Satellite Data", IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-11, no.4, pp. 312-318.

Barr, A., Feigenbaum, E.A., 1981, The Handbook of Artificial Intelligence - Vol. I, Pitman.

Winston, P.M., 1977, Artificial Intelligence, Addison-Wesley, Reading Mass.

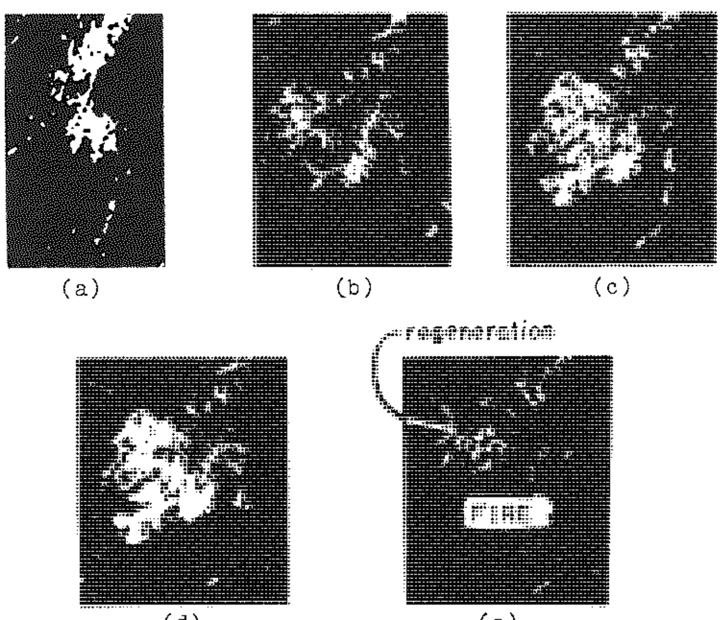


Figure 3. Successive Decisions Indicating the Progression of Logging

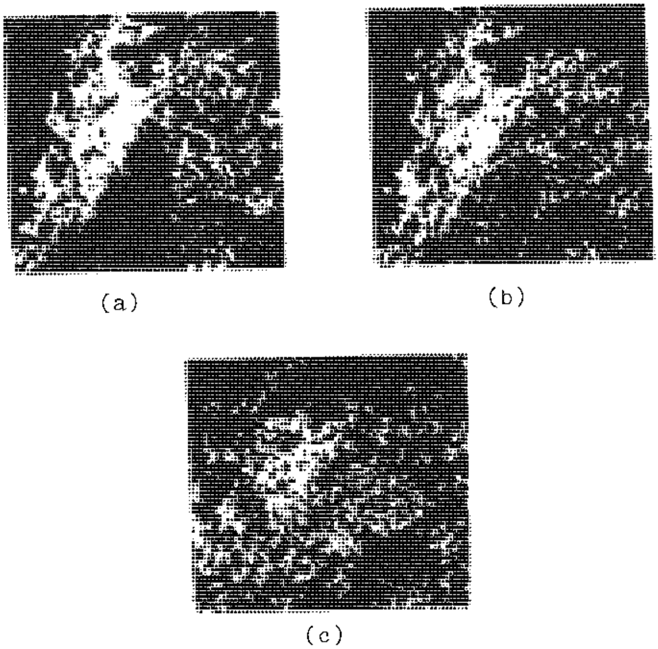


Figure 4. Successive Decisions Indicating the Progression of Defoliation due to Spruce Budworm Infestation

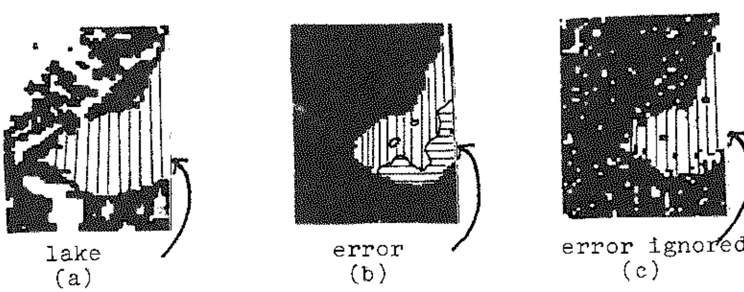


Figure 5. Classification Correction by the FORESTRY EXPERT PACKAGE

LANDSAT DIGITAL DATA FOR UPDATING GLACIOLOGICAL INFORMATION ON TOPOGRAPHIC AND GLACIER INVENTORY MAPS

Philip J. Howarth
Department of Geography, McMaster University
1280 Main Street West, Hamilton, Ontario, Canada L8S 4K1
and
C. Simon L. Ommanney
Snow and Ice Division, National Hydrology Research Institute
Inland Waters Directorate, Environment Canada
Ottawa, Ontario, Canada K1A 0E7

ABSTRACT

Landsat provides repetitive coverage of the earth's surface. The technical characteristics of its multispectral scanner make the sensor best suited to the study of large areas at relatively low resolution where there are clear spectral and/or spatial changes which are either permanent or long term. Many changes important for map revision meet the above criteria. In Canada, Landsat visual data are incorporated in the topographic map revision process, not only for mapping change but also in the planning of new aerial photography for revision mapping. In this paper, the need for updating glaciological information on topographic and glacier inventory maps is discussed and the capabilities of Landsat digital data to undertake this task are demonstrated with three examples. On the Steacie Ice Cap, a 5 km advance of the Good Friday Bay Glacier was detected with Landsat. A linear contrast stretch enhancement was used in the mapping of the terminal zone of the Kaskawulsh Glacier and the change in the surface moraine pattern on the Tweedsmuir Glacier was also recorded. It is concluded that a digital data base using DICS imagery would provide easy and rapid updating of glacier margins for topographic and glacier inventory maps.

INTRODUCTION

For over a decade, the multispectral scanner (MSS) system on Landsat has been recording imagery of the Canadian environment. Given the size of the country and the sparseness of the population in many areas, satellite sensing is the only feasible method for acquiring up-to-date information about large, remote portions of the country. The fact that Landsat provides repetitive coverage of the earth's surface means that the satellite imagery has the potential to be an excellent source of information for detecting, recording and monitoring major changes in the landscape (Thompson et al. 1982). The aim of this paper is to demonstrate the capabilities of Landsat data, in digital format, for identifying glaciological changes that can be used to update topographic and glacier inventory maps.

LANDSAT AND ENVIRONMENTAL CHANGE

It is important to appreciate that the technical characteristics of Landsat limit the types of change that can be effectively recorded and analyzed using the satellite data. There are several factors to be considered.

The first factor is that Landsat is most suited to large area coverage at relatively low resolution. The pixel size or minimum ground area recorded by the MSS system is 57 m × 79 m. Thus, if changes are to be detected and recorded, at a minimum they should involve ground areas that cover several pixels and preferably the changes should be observable over relatively large lengths or areas.

A second factor to be considered is spectral and spatial change. Most environmental change involves a spectral change which is readily recorded by the MSS system. Any spatial changes clearly have to be of the order of several pixels (i.e. a few hundred metres) to be observable on the Landsat imagery.

The final factor involves time. Allowing for problems of cloud cover on many satellite passes and the 16-day repeat cycle of Landsat 4 (18 days with Landsats 1-3), the changes that are best studied involve either a long-term or a permanent change (Howarth et al. 1982). Although the Landsat satellites have only been recording data for just over a decade, the imagery can be compared with older maps or aerial photographs to determine changes over a longer period of time.

Thus it can be seen that for recording and monitoring changes, Landsat is best suited to the study of large areas at relatively low resolution where there are clear spectral and/or spatial changes and where the changes are either permanent or long term. From a theoretical point of view, the study of glaciological change for updating topographic or glacier inventory maps clearly meets the above criteria.

PREVIOUS STUDIES

The idea of using Landsat imagery for map revision has been considered for some time. Pioneering work in Canada by the Topographical Survey Division of Energy, Mines and Resources and by Gregory Geoscience Ltd. of Ottawa has now taken its use to the practical stage where Landsat imagery is incorporated into various aspects of the revision cycles for 1:250,000 and 1:50,000 scale maps of the National Topographic System (NTS).

The imagery can be used for several purposes. First, verification of the correct location of physical features can be carried out using a combination of visual and digital data. This is of particular value where isolated islands in the coastal zone or within large lakes cannot be accurately located using standard photogrammetic methods (Fleming 1977; Fleming and Lelievre 1977). Geometrically-corrected data registered to the Universal Transverse Mercator (UTM) projection, as produced by the Digital Image Correction System (DICS) at the Canada Centre for Remote Sensing (CCRS), are particularly appropriate for this work (Fleming and Guertin 1980).

A second use of the Landsat imagery is for change detection. Fleming (1980) has reported on a study for a sparsely-populated area in which four NTS maps at 1:250,000 and 54 NTS maps at 1:50,000 scale were revised using both traditional methods and visual interpretations of optically-projected Landsat imagery (Gregory et al. 1982). All man-made changes significant for 1:250,000 scale maps were detected and virtually all changes for the 1:50,000 scale maps, although in the latter case identification of the nature of the change was not always possible. Accuracy of positioning was significant for revision overprinting (Fleming 1980).

A third use of Landsat imagery, related to the above study, is to identify areas that do not need flying of new photography for revision purposes. Fleming (1980) deduced that only 10% of the study area would have required new photography if Landsat had been used in the change detection process.

Revision cycles for the update of topographic maps have been established for different areas in Canada. The maximum length of time between revisions is ideally 30 years for remote or arctic and subarctic areas, although in recent years it has not been possible to maintain this cycle. Fleming (1981, p. 94) indicates that the changes in remote areas will "tend to be limited to the development of settlements and the construction of roads, pipelines, power lines or reservoirs". The authors of this paper would contend, however, that changes in the ice margins of glaciers can also be considerable and should be included in the revision process.

METHODOLOGY

The map revision reported by Topographic Surveys and Gregory Geoscience Ltd. has primarily involved visual analysis of the Landsat data. Given the large areas to be covered in map revision and the enlargement and overlay capabilities of the PROCOM system (Gregory et al. 1982), this is a valuable and cost-effective approach. In the present study, however, a digital approach was used to ensure that the maximum spectral and spatial resolution could be extracted from the data. In addition, extraction of glaciological information for map update formed only part of the investigations.

The analysis was carried out using the image analysis system (CIAS) at the Canada Centre for Remote Sensing (CCRS). The characteristics and capabilities of the CIAS have been described by Goodenough (1979). Particular advantages of the system for this study are its good display capabilities (24 bit image) and a large electrostatic printer attached to the CIAS that is capable of producing plots at any scale.

STUDY AREAS

The capabilities of Landsat digital data for providing glaciological information for updating topographic and glacier inventory maps are demonstrated with three examples.

Steacie Ice Cap

The Steacie Ice Cap is located at the southwestern corner of Axel Heiberg Island in the high Arctic. The ice cap, with almost 200 outlet glaciers, covers an area of approximately 3,000 km^2 (Ommanney 1969). It is typical of the polar ice caps and cold-based glaciers, of much of the high Arctic. The 1:250,000 NTS map of the area (Glacier Bay, 59E) is based on aerial photographs taken in 1958-1960 and was printed by the Topographical Survey Division in 1966. The same aerial photographs were used to produce the glacier inventory of the ice cap, including the maps of the area (Ommanney 1969; 1970).

The Landsat image selected for study was 20948-18430 (+ 10 secs.), recorded on 27 August 1977. It represented typical ground conditions for the end of the ablation season when ice margins and permanent snow patches would not be obscurred by seasonal snow. The full image was displayed on the colour monitor of the CIAS and a subscene was selected to display the whole ice cap. It consisted of 1107 lines of data with 1536 pixels per line. The image was, therefore, approximately geometrically correct when viewed on the colour monitor of the CIAS and represented a ground area of 87.5 km of 87.5 km.

The end of August is more than two months after the summer solstice and at a latitude of 78°30' the sun altitude was only 20° when the satellite passed over in the morning. Although this helped to show detail in the topography due to shadow, radiance from the terrain was low. A linear contrast stretch was applied to each of the four Landsat bands. This produced a brighter image giving detail in the terrain and emphasizing the boundaries of the glaciers.

From the digital image of the ice cap a 16-level, grey-tone print was generated at a scale of 1:125,000, the same scale as the original glacier inventory compilation. Overlaying the print with the glacier inventory compilation on a light table, it was possible to determine that the Good Friday Bay Glacier had advanced approximately 5 km since 1959, the year of the aerial photography of the glacier. There appeared to be no changes in position of the other ice edges. Figure 1 is a photograph of part of the grey-tone print showing the Good Friday Bay Glacier. For comparison, the same area traced from the 1:250,000 scale NTS map of the area is shown in Figure 2. The major differences in location of the ice edge between the two dates can readily be identified.

Of practical importance is the fact that the glacier now ends in the fjord and is, therefore, capable of producing icebergs. It is the only outlet glacier of the Steacie Ice Cap that currently produces icebergs. These could become a hazard to navigation in a previously iceberg-free area.

Kaskawulsh Glacier

Located at the eastern end of the St. Elias Mountains, the Kaskawulsh Glacier is one of the larger Canadian glaciers. The image selected for study was 21314-19293 (+ 10 secs.) which was recorded on 28 August 1978. An enlarged subscene measuring 184 × 256 pixels (14.5 km × 14.5 km) was selected to display the tongue of the glacier. A linear contrast stretch was applied to the image and a supervised classification was also produced.

The raw image was relatively dark and minimum and maximum digital values for the area displayed were Band 4 (3 and 20), Band 5 (3 and 23), Band 6 (2 and 22) and Band 7 (1 and 17). For each band the lower limit was set to 0 and the upper limit to 63. All the digital values were adjusted in a linear fashion between these limits, thereby extending the dynamic range of each band.

On the enhanced image, major features could be readily interpreted. Even though the tongue of the glacier was covered with moraines, the ice edge could be identified and the glacier differentiated from the proglacial area. This was done in a variety of ways. On one side of the ice front a more textured surface was produced by a pattern of lighter and darker pixels suggesting a hummocky surface to the glacier. In other parts, the ice edge was marked with a shadow while a proglacial lake in contact with the ice edge also helped to indicate the maximum extent of the glacier. In the proglacial area, sediment-laden water forming a braided channel system could be identified. An end moraine and the vegetation distribution could also be interpreted from the image.

An attempt was made to produce a supervised classification of the image. It was easy to separate ice from debris-covered ice and to identify the proglacial lake. However, differentiating debris-covered ice from parts of the proglacial area and valley-side slopes was not possible. Highlights and shadows in the mountains also caused considerable confusion in the classification.

For this fairly complex environment at the glacier margin, it was concluded that supervised classification is of little value. However, an enhanced image and an electrostatic printer output at a scale of 1:50,000 (Figure 3) could readily be interpreted.

Tweedsmuir Glacier

The final study was carried out on the Tweedsmuir Glacier using imagery from two different dates (13 September 1973 and 28 August 1978). The glacier surged in the early 1970s and nearly dammed up the Alsek River, which led to its early study (Post et al. 1975). In the present investigation, the aim was to detect changes in the patterns of the medial moraines on the ice surface.

The methodology involved registering the images from the two dates. The best method would have been to use the preprocessed DICS imagery (Butlin et al. 1978), but two problems arose. First, it proved impossible to read the original Canadian tape for the early date and imagery had to be acquired from the USA. Second, the lower half of the glacier is traversed by a UTM zone boundary making production of imagery difficult. For these reasons, image to image registration was carried out.

Combinations of Bands 5 and 6 from the two dates were displayed on the monitor of the CIAS. From the colour patterns that were produced, it was possible to identify which sections of the medial moraines had moved between the two dates. In other words, this experiment demonstrated that changes in pattern of surface moraines on glaciers can be readily identified and then mapped.

DISCUSSION

Although the majority of glaciers are in the more remote areas of the country, there are often compelling reasons, in addition to the desire for accuracy, to have the location of ice margins updated on topographic maps. The example has been given for the Good Friday Bay Glacier where icebergs are now being produced in a previously iceberg-free environment. The presence of the icebergs would have an effect on offshore activities in adjacent areas. Where there is current or potential economic activity in remote areas, the revision cycle is likely to be more frequent and it is important to have up-to-date maps for planning purposes. In some circumstances glaciers can create hazards (particularly if they surge) and information on the positions of such glacier margins is important. Pipeline routing, for example, could be influenced by the presence or absence of a glacier margin.

Another reason for updating glacial information is that inaccurate location of glacier margins on topographic maps can lead to confusion. An example occurred several years ago with the Grand Pacific Glacier. The glacier flows towards Tarr Inlet at the head of Glacier Bay, Alaska. Both the 1952 aeronautical chart and the 1962 1:250,000 scale NTS map (114P) show the 1925 position of the ice margin. At that time there was land on the Canadian side of the border with the USA, which crosses the area. This led to the suggestion that a Canadian port could be constructed to provide access to the Yukon. In reality, the glacier margin was much further advanced and did in fact end in the waters of Tarr Inlet. This made any port construction impossible.

In this paper, the cost effectiveness of a visual versus a digital approach has not been considered. Gregory et al. (1982) have given figures which indicate that for large areas a visual approach would be difficult to beat for cost effectiveness. For glacial areas, however, we are dealing with much more limited areas of the country. If a data base consisting of DICS imagery were to be established for the major glacial areas, it would be possible to rapidly compare new imagery with the data base and flag any significant changes. An enhancement approach to identifying the changes, as illustrated in Thompson et al. (1982), could readily be adapted to this task. In addition to being used for updating topographic maps, the data base could also be used for glacier inventory purposes and perhaps in other glaciological studies (Howarth, 1983).

If changes in glacier margins are identified, the question is how to show them on topographic maps. New mapping of relevant areas from aerial photographs could be an expensive procedure, particularly if glacier fluctuations were the only changes on the map. Overprinting would thus appear to be the answer to show both advances and retreats of ice margins. Clearly there would be a change in contours as well, but these could not be obtained from the satellite data. If an ice retreat occurred, the morphology of the recently-exposed landscape would not be depicted on the map. However, the person using the map would be able to deduce and understand the nature of the changes that had occurred and, more importantly, would have an up-to-date topographic map for the area being studied.

CONCLUSIONS

Both from theoretical and practical points of view, Landsat data are suited to updating glaciological information on topographic and glacier inventory maps. The digital data provide maximum resolution and permit enhancement to aid in detecting and mapping changes. The need for such information in areas of current or potential economic activity was demonstrated.

Although a visual approach is more cost effective for total area coverage, it is suggested that a digital data base using DICS imagery might be appropriate for identifying and mapping changes in the more limited areas covered by glaciers. Rapid output of geometrically-corrected electrostatic printer plots would show changes and this information could be included as an overprint on a revised topographic map.

ACKNOWLEDGEMENTS

This paper is based on work that was undertaken by Dr. P.J. Howarth for the National Hydrology Research Institute of Environment Canada. Financial support for the study is gratefully acknowledged. The digital analysis would not have been possible without the willing co-operation of the Canada Centre for Remote Sensing in permitting access to their Image Analysis System (CIAS). The advice and assistance of Mr. Bill Bruce of CCRS as scientific liaison for the study is gratefully acknowledged. Operator assistance on the CIAS was ably provided by Al Banner and Mike Manore.

REFERENCES

Butlin, T.J., F.E. Guertin and S.S. Vishnubhatla 1978, The CCRS Digital Image Correction System: Proceedings of the 5th Canadian Symposium on Remote Sensing, Victoria, B.C., pp. 271-276

Fleming, E.A. 1977, Positioning off-shore features with the aid of Landsat imagery: Photogrammetric Engineering and Remote Sensing, Vol. 43, No. 1, pp. 53-59

Fleming, E.A. 1980, Change detection by Landsat as a guide to planning aerial photography for revision mapping: Proceedings of the 14th Congress of the International Society for Photogrammetry, Hamburg, West Germany, 11 pp.

Fleming, E. 1981, Monitoring revision requirements for Canadian maps: Proceedings of the 7th Canadian Symposium on Remote Sensing, Winnipeg, Manitoba, pp. 93-99

Fleming, E.A. and F.E. Guertin 1980, Determination of the geographical position of isolated islands using the Digital Image Correction System for Landsat MSS imagery: Proceedings of the 14th Congress of the International Society for Photogrammetry, Hamburg, West Germany, 10 pp.

Fleming, E.A. and D.D. Lelievre 1977, The use of Landsat imagery to locate uncharted coastal features on the Labrador coast: Proceedings of the 11th International Symposium on Remote Sensing of Environment, Ann Arbor, MI, pp. 775-781

Goodenough, D.G. 1979, The Image Analysis System (CIAS) at the Canada Centre for Remote Sensing: Canadian Journal of Remote Sensing, Vol. 5, No. 1, pp. 3-17

Gregory, A.F., H.D. Moore and A.M. Turner 1982, PROCOM: a cerebral processing system for analysis of Landsat and collateral data at scales up to 1:15,000: Proceedings of a Symposium on Advances in Instrumentation for Processing and Analysis of Photogrammetric and Remotely Sensed Data, International Archives of Photogrammety, Vol. 24-11, pp. 237-247

Howarth, P.J. 1983, Evaluation of Landsat digital data for providing glaciological information: Contract Report for National Hydrology Research Institute, Environment Canada, Ottawa, 43 pp.

Howarth, P.J., T.T. Alföldi, P. Laframboise, J.C., Munday, Jr., K.P.B. Thomson, G.F. Tomlins and G.M. Wickware 1982, Landsat for monitoring hydrologic and coastal change in Canada: in Thompson, M.D. et al. (eds.) Landsat for Monitoring the Changing Geography of Canada, Canada Centre for Remote Sensing, Energy Mines and Resources, Ottawa, pp. 7-40

Ommanney, C.S.L. 1969, Glacier Inventory of Canada: Axel Heiberg Island, Northwest Territories: Technical Bulletin No. 37, Inland Waters Branch, Department of Energy Mines and Resources, Ottawa, 98 pp.

Ommanney, C.S.L. 1970, Glacier Inventory of Steacie Ice Cap Area (79°N, 90°W), Axel Heiberg Island, N.W.T.: Reprint Series No. 65, Inland Waters Branch, Department of Energy Mines and Resources, Ottawa, 35 pp.

Post, A., M.F. Meier and L.R. Mayo 1976, Measuring the motion of the Lowell and Tweedsmuir surging glaciers of British Columbia, Canada: in Williams, R.S., Jr. and W.D. Carter (eds.) ERTS-1: A New Window on our Planet, United States Geological Survey, Professional Paper 929, pp. 180-184

Thompson, M.D., P.J. Howarth, R.A. Ryerson and F.J. Bonn 1982, eds: Landsat for Monitoring the Changing Geography of Canada, Canada Centre for Remote Sensing, Energy Mines and Resources, Ottawa

Figure 1. Part of an electrostatic printer output of the Steacie Ice Cap showing the Good Friday Bay Glacier (upper centre) in August, 1977. The original scale of the output was 1:125,000.

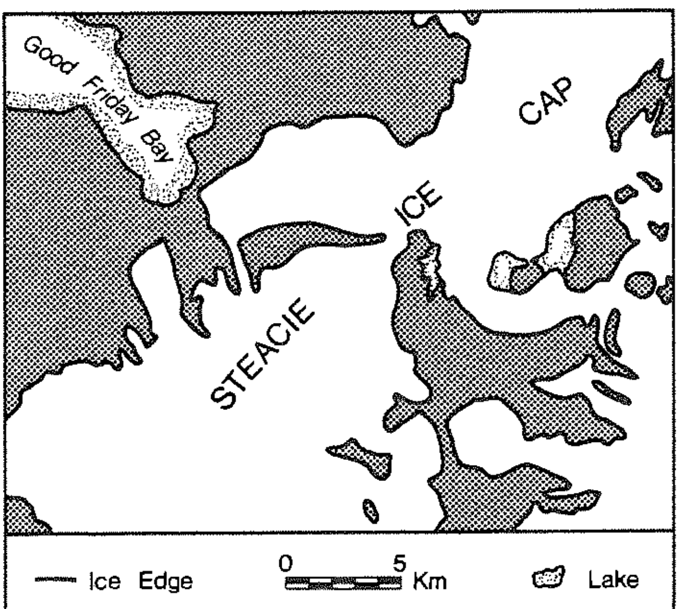


Figure 2. The same area as shown in Figure 1, but traced from the 1:250,000 scale NTS map printed in 1966. Comparison of the two figures clearly shows changes in the Good Friday Bay Glacier.

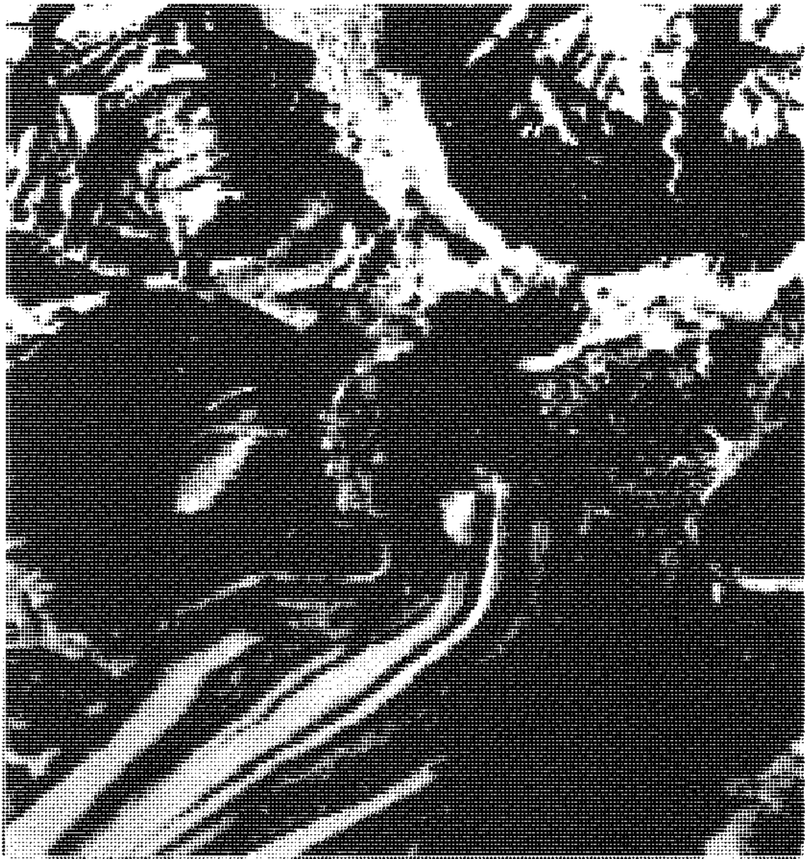


Figure 3. A 16-level, grey-tone plot from an electrostatic printer (Versatec) showing the terminal zone and proglacial area of the Kaskawulsh Glacier. Scale of the original plot was 1:50,000.

ENHANCEMENTS AND CLASSIFICATIONS OF LANDSAT DATA FOR THE ECOLOGICAL RESOURCE SURVEY AND MAPPING OF THE AISHIHIK LAKE ENVIRONMENTALLY SIGNIFICANT AREA (ESA) YUKON TERRITORY, CANADA

D. Puittinen and A.B. Kesik
Department of Geography, University of Waterloo
Waterloo, Ontario, N21 3G1, Canada

ABSTRACT

Enhancements and classifications of the Landsat digital data for the Aishihik Lake Environmentally Significant Area (ESA) have been analysed as the supporting techniques for the biotic survey and mapping. Digital techniques provide opportunity for dynamic manipulation of data and for more effective interactive interpretation. Enhancement techniques, particularly linear contrast stretch, seems to be the most useful for the objectives and requirements of the ABC resource survey.

INTRODUCTION

Resource survey and mapping in unaccessible areas depend strongly on remote sensing data and their analysis. Diversified remote sensing data, thanks to the addition of new platform/sensor systems, stimulated efforts for the integration of multisensor data with the more conventional systems of information (Estes, 1982). This approach requires precise analysis of properties and characteristics of each set of data, as well as quantitative techniques of data manipulation. Whether or not integration is suitable for a given application may be decided consequently.

The global Landsat System provides data which may be converted into useful information for multi directional environmental studies, thematic mapping and terrain resources management (Townshend, 1981). Most effective transformation of spectral, spatial and temporal Landsat data into information is based on the digital techniques of image enhancement and classification (Swain and Davis, 1978).

The enhancement techniques attempt to maximize the visual information presented on the CRT monitor. These techniques may stimulate analysts for more efficient mental, interactive interpretation. Such an interpretation involves search for correlations and interrelations between the visually displayed properties of Landsat data and other characteristics of environments represented or by environmental knowledge of analysts or by other accessible sources of environmental data.

Landsat classification techniques with their two fundamental mutations; unsupervised and supervised, as well as refinements offered by the contextual analysis, relay on the computer identification of spectral categories of terrestrial features (mostly integrated) interpreted and labeled as "thematic" classes. Majority of efforts in the Landsat image classification have been oriented towards documentation of the usefulness of classifications for particular "mapping" projects dealing with the natural or man-made components of environments (Johannsen and Sanders eds. 1982).

Thirty-five sites, representing 25% of the Yukon Territory have been proposed as the Environmentally Significant Areas (ESA) (Theberge et al., 1980). These sites require environmental survey with emphasis put on the inventory and assessment of abiotic (A), biotic (B), and cultural (C), or ABC components of environments. The concept of the ABC survey and mapping has been elaborated by environmentalists; Bastedo, Nelson, Theberge (Bastedo, 1982). It involves multi-level survey and mapping of the ABC components from the preliminary inventory at the level I to the management oriented synthesis at level IV.

The ABC method is viewed as an alternative to the ecological land classification approach, promoted in Canada by the Land Directorate of the Department of Environment. The ABC alternative puts more emphasis on the parallel multi-disciplinary survey of all three components with a hybrid methodology of surveying and mapping (Bastedo, 1982). The major thrust of the ABC resource survey is the mapping of the "structural" and "functional" characteristics (variables) of the ABC components. Structural variable characterize the properties and conditions of the ABC components of given ESA, eg. morphometry of landforms (A), species composition (B), burial grounds (C). Functional variables could include processes associated with the origin and evolution, eg. thermokarst activity (A), uniqueness of habitat (B), present and historical land use (C).

The ABC ecological resource survey has been applied by ecologists to the Aihihik Lake ESA in the Yukon Territory comprising 128,865 km^2 (Fig. 1). The Aishihik Lake ESA is located in the Alsek Drainage Basin (Pacific Basin) and it is represented mostly by the alpine and sub-alpine environments. Local relative elevation reach 1220m. ESA territory comprises sections of four physiographic units: Kluane Plateau, Aishihik Basin, Ruby Range and Nisby Range (Tempelman-Kluit, 1974).

From being in progress ABC survey and mapping, completed survey refers to the biotic (B) component investigated by Bastedo (Bastedo, 1982). He compiled Biotic Land Units (BLU), structural map in the scale of 1:250,000 using landscape approach, and the map of the "Spacial Habitat Zones", representing synthesis of the functional variables. The methodology of BLU mapping relies strongly on the application of photo interpretation of conventional B&W aerial photographs and the visual-manual interpretation of Landsat images represented by the 20x20 cm colour composite transparencies analysed with the aid of Zoom-Transfer Scope (ZTS). The essence of application of Landsat images may be summarized as follows:

Prefield Stage Application:

- Visual interpretation of Landsat colour transparencies and recognition of Landsat "themes" (approx. 10). "Themes" represent image areas with similar interpretative indicators. They are subsequently identified as preliminary BLU's.

- Plotting of recognized preliminary BLU's into a base map in the scale 1:125,000. Minimal size of the unit, 2.5x1 cm or corresponding field area of 6 km^2.

- Identification of areas or zones difficult for thematic classification to the BLU's and designated for the subsequent field study.

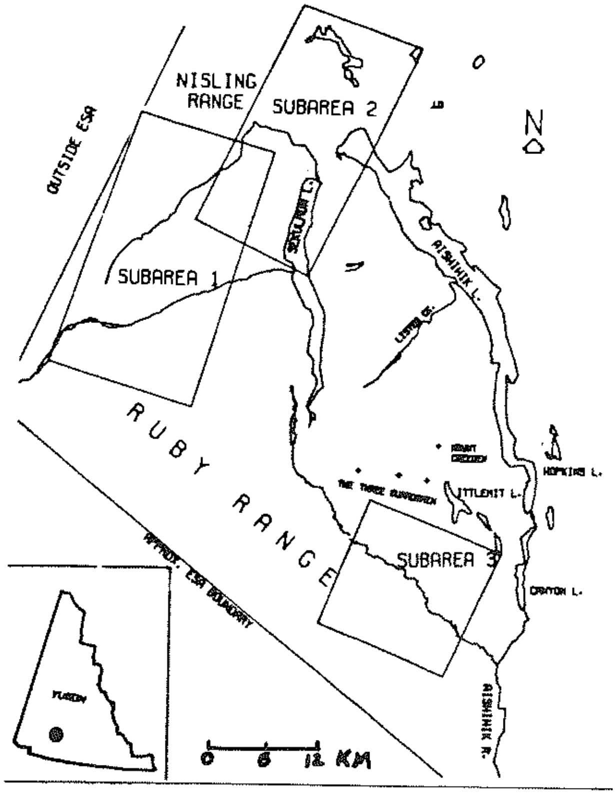


Fig. 1. Aishihik Lake ESA with subareas selected for the supervised classification training and assessment of Biotic Land Units delineation.

Field Work Applications

- Matching of Landsat "themes" recognized as preliminary BLU's with the vegetation pattern in the field.

- Using of Landsat imagery for selection of sampling areas.

The visual-manual techniques of Landsat imagery analysis applied by the ABC resource survey researchers represent a simplistic approach which do not explore the dynamic nature of the Landsat digital data. Beyond any doubt the simplicity of visual-manual techniques and the similarity to the conventional photo interpretation techniques play an important role at the introductory level of Landsat data interpretation. However, it is also well understood that any new type of resource survey should investigate the most effective and economic applications of Landsat data for specific, user oriented application. Such an approach implicates the examination of existing Landsat image enhancement and classification techniques and their assessment for the resource survey objectives. In this respect presented research may be categorized as a contribution to the study on methodological alternatives and complementary applications of Landsat imagery for the ABC biotic resource survey and mapping.

LANDSAT DATA AND IMAGE ANALYSIS TECHNIQUES

A 185 by 185 kilometers Landsat scene can be acquired from the Canada Centre for Remote Sensing (CCRS) in several forms. Commonly used for vegetation mapping is a false-colour 1:1 million transparency. Such an image was the Landsat product used as an integral part of the ESA ABC survey for preliminary mapping of Biotic Land Units. Colour and textural variations across the image corresponding to changes in vegetation pattern and physiography were interpreted to denote the land units. Landsat makes use of a multispectral scanner (MSS) to sense electromagnetic radiation in four bands: band 4 measuring reflected radiation between 500 and 600 nanometers (blue-green) band 5 between 600 and 700 nanometers (red), band 6 between 700 and 800 nonometers (near infrared) and lastly band 7 between 800 and 1100 nanometers (near infrared). The false-colour composite image used for Biotic Land Units mapping results from the sequential exposure of band 4 through a blue filter, band 5 through a green filter and band 7 through a red filter.

By employing the capabilities of a Dipix Image Analysis System, far greater information can be extracted about the Landsat scene with it in the form of a digital image than can be interpreted from the static and necessarily degraded false-colour image. Acquired on a computer compatible tape (CCT) from CCRS, the scene consists of all four MSS bands. Each band is comprised of 2286 scan lines, 3600 pixels across. Each pixel has a ground resolution size of 57 metres by 79 meters. The brightness value recorded for each pixel represents an integration of the brightness of all landscape features within the IFOV.

The Landsat CCT scene used in this study was of latitude centre 61.21 degrees north and of longitude centre 137.03 degrees west. An area of 1400 pixels by 1250 lines was extracted for the analysis as an image of that size was determined sufficient to contain the Aihihik Lake ESA. The image was sensed by the Landsat 2 satellite on August 28, 1978 and was cloud-free. No other good images were available of the area.

Digital processing of the CCT Landsat data included correction of the systematic distortions attributable to mirror velocity variability and Earth rotation. DICS tapes were not considered as the Dipex Image Analysis System possesses its own geometric correction routine. Geometric correction of the Landsat imagery was performed to the UTM reference system by identifying control points on 1:250,000 NTS topographic maps.

Field and Accessory Data

A purposive method (Townshend, 1982) of field sampling was employed for the study. Three subareas (Fig. 1) representing a subalpine, an alpine, and a major river valley environment, recognized to contain all the major physionomic vegetation communities in the ESA were identified. Each was field checked according to its accessibility. The alpine environment required helicopter transportation.

The objective of the field investigation was not necessarily to identify training and test sites for the Landsat analysis procedure but to determine or verify the identity of the land covers interpreted on aerial photographs. Full stereo aerial photograph coverage, 1:63,360 scale, of the Aishihik ESA was obtained and used as the major training and testing data source for the analysis procedures.

Enhancement Techniques

Using routines developed for the Dipix Image Analysis System, the enhancement methods of linear contrast stretch, principal component analysis and band ratioing were applied to the Aishihik Lake ESA digital Landsat image. From an array of enhancement products the most useful to ESA map compilation was then judged based on the interpretability of the major physiognomic vegetation coverings in the ESA.

Contrast stretch enhancement can be obtained automatically to the full brightness range (0-255) of the Dipix monitor display. A false-colour composite of band 4-blue, band 5-green and band 7-red, i.e. the same composite colour scheme as the preprocessed transparency image, was produced (Fig. 2). From the image ten major physiognomic vegetation communities, determined by ecologists, can be easily recognized by their distinctive interpretative indicators, mostly colours. The reduction of the image to encompass just the ESA region restricted the contrast stretch to the pixel brightness range contained in the ESA. This resulted in a greater stretch than would have occurred for the entire Landsat scene which contained very high pixel brightness values in all bands corresponding to high reflectance from ice fields in the St. Elias Mountains. The colour contrast on the Landsat 1:1 million false-colour transparency is poor for this reason.

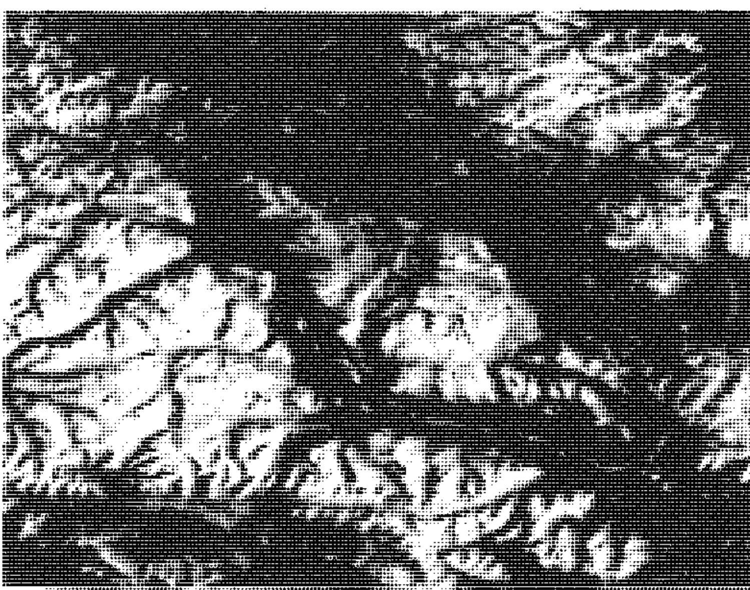


Fig. 2. Digitally enhanced false-colour image.

The analysis for creating principal component images was developed by Taylor (1974). Produced using the analysis are four eigenvectors or new uncorrelated band dimensions to which the Landsat image data are redescribed. The eigenvectors are first generated from a training area containing a land cover variation representative of the whole ESA. If, as opposed to the general land cover enhancement wanted in this study, enhancement of a particular land cover was desired, the training area land cover would be weighted in favour of that cover. For display on the colour monitor the first component image which accentuates the relief variations in the ESA emphasizing shadows and sun facing slopes is used to control brightness. The second component image providing good discrimination of high reflectance regions like the high alpine, eroded valley slopes and beach areas is used to control a red-green hue axis. Incorporated as a second hue variation axis, blue-yellow, is the third component which gives the best differentiation of lakes in ESA. This form of image composition permits rotation of the hue axes where the opposite ends of the rotation would be dominated by red-yellow and blue-green colourations. For the Aishihik ESA the non-rotated principal component image was the most interpretable. However, when compared to the false-colour contrast stretch image, it was found less useful for mapping the major physiognomic vegetation coverings, and thus has less value to ESA survey. An image comprised with the first component replacing band 4 was additionally produced and although it did provide better enhancement of certain features, it was still poorer for overall interpretation of the ESA vegetation cover. The first component image is, in fact, very similar to band 7.

A new image created by ratioing band 7 over band 5 was incorporated in a composite along with bands 4 and 7 to examine the possible benefits of the ratio enhancement. Thus instead of two highly correlated visible bands and one infrared band the composite was comprised of a visible, an infrared, an infrared/visible ratio band. Four broad land covers were better differentiated, based on more distinctive colouration.

In conclusion, a Dipix digital Landsat image displayed in the form of a full range linear contrast stretch false-colour image (Fig. 2) is vastly superior to the standard false-colour transparency for the objectives of discussed resource survey. Creation of new bands via principal components analysis or band ratioing, on the other hand, does not result in a further increase in the interpretability of the image.

Classifications

The two classifications of Landsat scene; unsupervised and supervised have been applied to the Aishihik image. Unsupervised classification involves identification of spectral classes based on statistical analysis of the four band histograms. Thirteen classes was considered an appropriate number for dividing the area up into its physiognomic vegetation classes. The thirty-two most significant histogram maximas were first identified and described by unimodal distribution for each of the four bands. The thirty-two maxima were next reduced to thirteen classes based on overlap between distributions. The image was then maximum likelihood classified by assigning each pixel to the distribution or class to which it has the highest probability of belonging. By combining poorly separated classes, a well differentiated image of ten classes corresponding to defineable physiognomic vegetation covers was produced.

In supervised classification the class distributions are defined by training samples, areas that have been field checked and which represent land covers desired for mapping. The Aishihik Lake ESA was divided into ten generalised physiognomic vegetation covers using supervised classification, (Fig. 3).

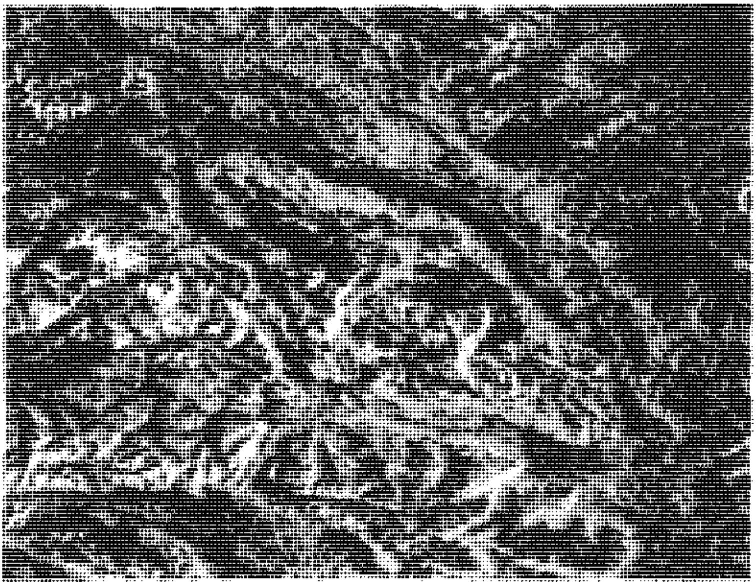


Fig. 3. Supervised classification of Landsat image.

Table 1: Legend for Supervised Aishihik Lake ESA Classification

CLASS	COLOUR	PIXELS	%SCENE
water	blue	57152	3.26
woodland	green	463090	26.46
forest	dark green	113586	6.49
alpine tundra	magenta	181371	10.36
alpine barrens	cyan	23126	1.32
upland shrub	yellow	150897	8.62
shrubland	red	534365	30.53
alpine shrub	pink	133966	7.66
eroded slope	buff pink	24563	1.40
shadow	dark blue	63278	3.61
unclassified	white	4606	0.26

Cartographic display of the supervised classification with geometric correction has been presented in the form of Applicon colour map produced at OCRS.

A comparison of the unsupervised and supervised classifications demonstrates the advantages of precise specification of the class ranges derived in supervised classification from the training samples. Unsupervised classification relies on spectral separation between classes. In the Aishihik Lake ESA the physiognomic vegetation covers are characterized by a gradual transition. This resulted in, for example, a class better described as woodland-shrubland than either woodland or shrubland.

Dipix displayed contrast stretch false-colour image allows for easier and more effective delineation of biotic land units. Fig. 4 illustrates the subarea 2 presented as the contrast stretched and unsupervised images (Fig. 4).

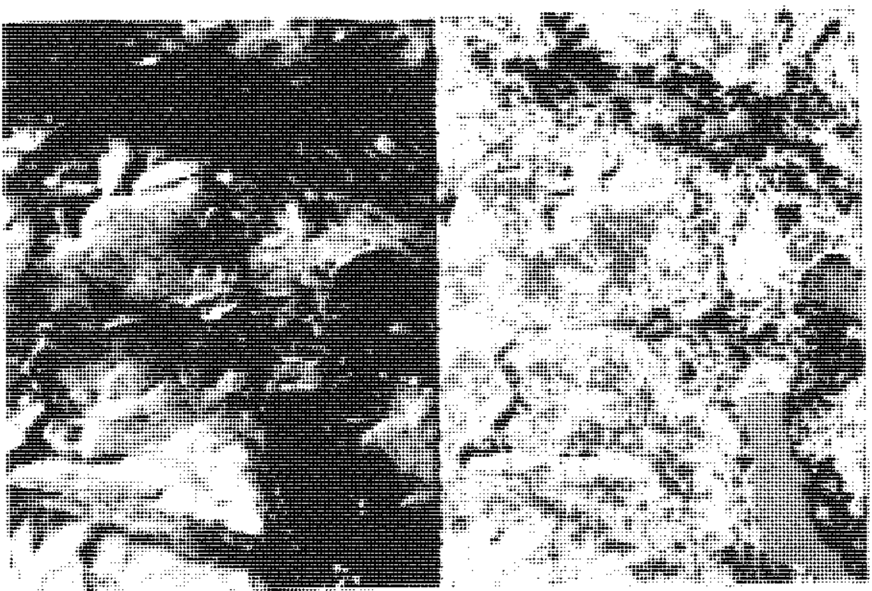


Fig. 4. Subarea 2. False colour contrast stretch and unsupervised classification.

Both images have been rectified as best possible to a 1:250,000 NTS map using the rotation, stretch and magnification capabilities of the zoom transfer scope (ZTS). Land units were then delimited on the two Landsat products and compared with those derived from the biotic surveying with the aid of 1:1 million Landsat transparencies (Fig. 5). The subarea was small enough for the ZTS rectification. Rectification of the entire ESA would require division of the image into sections. Therefore geometric correction using the Dipix system is recommended.

On the false colour image the most interpretable information that can be used for depicting land units refers to the physiographic details such as valley and ridges. Conversely, the land units on the unsupervised classification must be derived from the heterogeneous colour pattern identification. For the use in ESA ABC resource survey the false colour image and physiographic approach to land unit identification should be preferred as it is more replicable.

In addition to delineating the biotic land units, digital Landsat analysis can be of value in rating the significance of biotic land units. Digital Landsat classifications and the possibility of overlaying land unit polygon boundaries and retrieving class distribution of the pixels contained in the units could greatly expedite the calculation of the indices of significance.

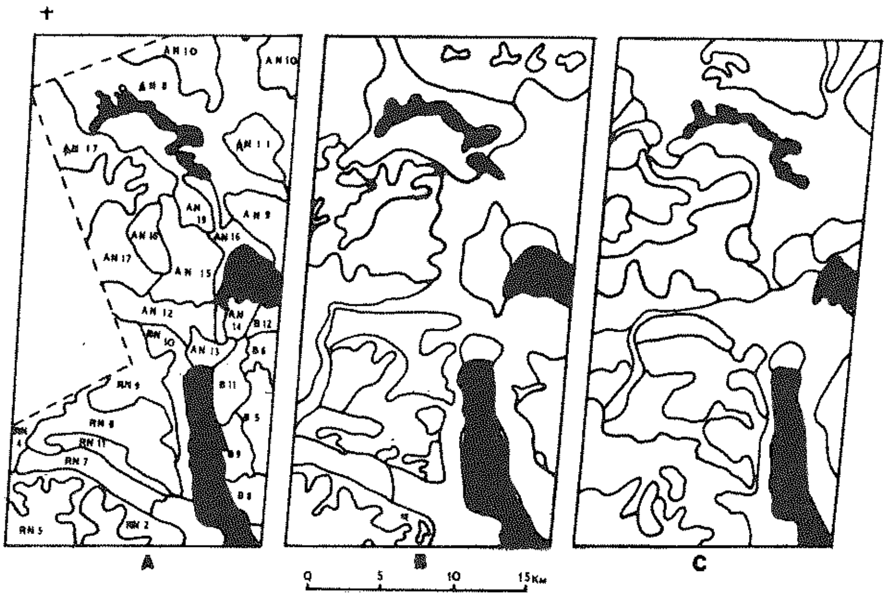


Fig. 5. A: Biotic Land Units after Bastedo (1982), B: Interpretative land units derived from the Dipix false colour contrast stretch display, C: Interpretative land units derived from unsupervised classification of Landsat image.

REFERENCES

Bastedo, J. A Resource Survey Method for Parks and Related Reserves with Special Reference to Environmentally Significant Areas in the Yukon. Dept. of Geography, University of Waterloo. (in press)

Estes, J.E. 1982. Remote Sensing and Geographic Information System coming of age in the Eighties. Proceedings, Pecora VII Symposium. Am. Soc. of Photogram. pp. 23-40.

Johannsen, Ch. J. and J.L. Sanders, 1982, eds. Remote Sensing for Resource Management, Soil Conservation Soc. of America.

Swain, P.H. and S.M. Davis, 1978, eds. Remote Sensing; The Quantittative Approach. McGraw-Hill Inc.

Taylor, M.M. 1974. Principal Component Colour Display of ERTS Imagery. Second Canadian Symp. on Remote Sensing, Guelph.

Tempelman-Kluit, D. 1974. Reconnaissance Geology of Aishihik Lake, Snag and Part of Stewart River Map Areas, West-Central Yukon. Geol. Surv. Can. Paper 73-41.

Theberge, J.B., Nelson, J.G. and T. Fenge, 1980. Environmentally Significant Areas of the Yukon Territory. CRC Research Monograph No. 4. Ottawa.

Townshed, J.B.G. 1982 Terrain Analysis and Remote Sensing. George Allen & Unwin, London.

LA CARTOGRAPHIE THÉMATIQUE EN TÉLÉDÉTECTION

J.R. Bélanger
Commission géologique du Canada
601 rue Booth, Ottawa, K1A 0E8

RÉSUMÉ

La carte thématique produite à partir de données de télédétection présente certains problèmes au géographe puisque la discrimination spectrale entre les phénomènes ne varie pas nécessairement en fonction des thèmes que l'utilisateur souhaite traiter. Certains de ces problèmes sont inhérents à la télédétection ou dus à des causes techniques échappant au contrôle du cartographe; cependant d'autres difficultés peuvent se rattacher à l'approche ou aux méthodes utilisées pour extraire l'information voulue des données spectrales. Il est donc important en cartographie thématique non seulement de maîtriser les différentes techniques d'analyse du paysage spectral mais aussi d'expliquer à l'utilisateur l'origine, les avantages et les limites de la taxonomie spectrale et éviter ainsi une interprétation erronée des scènes de télédétection. On donne ici à titre d'exemple la cartographie de formations géologiques de surface de l'Île du Roi Guillaume dans les Territoires du Nord-Ouest et de la cartographie de la végétation de la région de Thetford Mines au Québec.

INTRODUCTION

La carte thématique sert à décrire le paysage géographique en analysant ou exprimant des relations spatiales existant entre les formes d'origines multiples dont les combinaisons offrent des paysages (Rimbert et Lengellé 1969). Les données de base servant à dresser la carte thématique proviennent de parcelles d'information qualitatiges et quantitatives obtenues en mesurant certains aspects du terrain (Tomlinson 1972, p. 36).

Lorsque l'information de base permet d'identifier précisément les données géographiques le cartographe a accès à de multiples méthodes et techniques pour l'analyse et la reproduction graphique de la carte thématique. Cependant lorsqu'il s'agit de carte thématique produite à partir de données incomplètes, telles que les signatures spectrales générées par les balayeurs multibandes (MSS) de LANDSAT, le choix des méthodes d'analyse devient beaucoup plus critique et le géographe doit souvent faire appel à des techniques détournées pour extraire l'information géographique désirée.

Ainsi, il existe en télédétection différentes approches et méthodes pour l'analyse des données de base ayant pour but de produire des cartes thématiques; certaines méthodes utilisent une approche qualitative et d'autres une approche quantitative. Bien que chacune des deux approches soit valable pour l'extraction de l'information voulue des données spectrales, les résultats sont souvent très différents (Hajic et Simonett 1976, p. 378), c'est pourquoi il est très important de comprendre non seulement les méthodes d'analyse mais aussi la nature des données de base et de comprendre la limite de la télédétection dans la cartographie thématique. Après une brève description des méthodes d'analyse d'images de télédétection, on proposera une approche cherchant à allier les avantages des différentes méthodes pour produire les cartes thématiques.

LES DONNÉES DE BASE

En télédétection les signatures spectrales constituent des données de base pour l'analyse d'un paysage. Théoriquement chaque objet possède sa signature spectrale propre telle qu'une empreinte digitale; cette signature spectrale est exprimée par une courbe spectrophotométrique ou courbe de réflectance. Pour être unique il faut cependant que la courbe de réflectance couvre une partie suffisante du spectre électromagnétique. Or, les contraintes techniques actuelles ne permettent pas d'enregistrer les signatures spectrales complètes des objects, c'est pourquoi il faut procéder à un échantillonnage des courbes de réflectance à intervalles définis et reconstituer les signatures à partir de ces données. À cette contrainte il faut ajouter certains problèmes de résolution et de bruit. En effet les scènes MSS de LANDSAT sont formées d'une multitude de signatures correspondant à des parcelles de terrain d'environ 57 x 79 m appelées pixels; par conséquent chaque pixel peut englober plusieurs phénomènes ayant des signatures différentes. De plus, les signatures peuvent être faussées par l'interférence de l'atmosphère (Lintz et Simonett 1976; Turner 1973), par une variation de l'éclairage (American Society of Photogrametry 1973) ou par des défauts mécaniques (Bélanger 1978).

L'impossibilité d'obtenir des signatures complètes des objets ainsi que les problèmes de résolution et d'interférence augmentent donc le risque de confusion entre les différents phénomènes et empêchent une classification précise des signatures en classe spectrale. De plus, la taxonomie pose aussi certains problèmes en télédétection puisqu'elle diffère des classifications communément utilisées pour les cartes thématiques. En effet, les capteurs enregistrent la signature de tous les objets dans leur champ de vision sans distinction au niveau de leur signification, ce qui a pour résultat de présenter une foule d'information se rattachant à une multitude de thèmes sur une même scène. Ainsi "les phénomènes physiques que l'on mesure...ne varient pas forcément en fonction des catégories que l'utilisateur souhaite distinguer compte tenue de sa discipline et de l'application qu'il a en vue. On peut même partler de néotaxonomie..." (David et collaborateurs 1977, p. 560), ou encore de télétaxonomie.

L'INTERPRÉTATION DES IMAGES

L'interprétation des images de télédétection peut se faire à partir d'une approche soit qualitative ou d'une approche quantitative. Bien que la frontière entre les deux approches soit imprécise et que les deux approches soient souvent complémentaies, il reste que le point de départ où le principe de base de chacune d'elles est différent.

L'approche qualitative

Dans l'approche qualitative, les signatures spectrales sont reproduites sous forme analogique, telle que sur la photographie aérienne conventionnelle, et l'interprétation se fait à partir des différentes textures (teintes) et structures des phénomènes apparaissant sur l'image. L'interprétation est rendue possible en présentant à l'analyste un paysage familier où il peut faire appel à sont expérience pour dépasser l'identification directe des signatures individuelles et se servir du contexte régional pour regrouper les classes spectrales selon la taxonomie décidée.

L'approche quantitative

Dans l'approche quantitative les signatures spectrales sont enregistrées sous forme digitale et on se sert d'algorythmes mathématiques pour regrouper les signatures en thèmes ou classes spectrales et analyser les différentes composantes du paysage. Le regroupement des signatures peut se faire soit à partir de thèmes définis à l'avance à l'aide des données de terrain (méthode supervisée) ou selon une discrimination essentiellement spectrale (méthode non supervisée).

La classification supervisée procède en trois étapes: on choisit un certain nombre de signatures spectrales (échantillons) représentatives des classes que l'on veut former, on identifie les caractéristiques de chacune des classes puis on regroupe les pixels de la région étudiée à partir des caractéristiques des échantillons. Les caractéristiques des classes peuvent être exprimées soit par des paramètres statistiques (méthode paramétrique, exemple: maximum de vraisemblance, Reeves 1977, Goodenough 1976) ou (si l'on considère des signatures comme des vecteurs à n dimensions) selon la position absolue dans l'espace des coordonnées spectrales (méthode non paramétrique, exemple: fonction discriminante, Steiner et coll. 1972; parallélépipède, Alfoldi 1978).

La classification non supervisée regroupe les pixels en classes spectrales distinctes en se basant sur la similitude des signatures sans se soucier au départ de la signification des classes; il s'agit donc d'identifier les nuages cohésifs sur le plan statistique (Alfoldi 1978), voir figure 3, et faire correspondre chaque nuage à un thème sur le terrain. Il existe plusieurs méthodes d'analyse de dispersion pour identifier les nuages statistiques, parmi les plus courants on retrouve les méthodes d'analyse en composantes principales (Braconne 1974, Taylor 1973, Fontanel 1976) et les méthodes adaptives (Brun 1974, Diday 1972).

Critique des méthodes

Les approches qualitatives et quantitatives offrent chacune des avantages dans l'interprétation des scènes spectrales mais elles comportent aussi certaines lacunes au niveau de l'identification des phénomènes. L'approche qualitative présente un paysage familier à l'interprète mais le caractère analogique des signatures spectrales prive l'interprète d'une foule d'analyses basées sur les mathématiques pouvant mettre en relief certaines structures de l'image et établir des relations précises entre les différentes composantes.

L'approche quantitative, d'autre part, permet une analyse précise des signatures spectrales basée sur les mathématiques mais présente certains problèmes de taxonomie. Les méthodes non supervisées fondent la taxonomie sur la similitude des signatures selon leur proximité dans l'espace; chaque nuage spectrale correspond à une classe distincte. Or, certains nuages peuvent être séparés de façon artificielle à cause de problèmes techniques, ou encore, un même nuage d'alure gaussienne peut renfermer plusieurs phénomènes distincts. Par exemple, une région forestière, peuplée de feuillus et de conifères et d'espèces mixtes peut former un noyau spectral cohésif d'allure gaussienne même si celui-ci est composé de plusieurs classes distinctes sur le terrain (voir figure 4).

Les méthodes non supervisées peuvent donner de bons résultats pour des pixels ayant un haut niveau d'appartenance aux noyaux principaux, cependant lorsque ces pixels s'éloignent des centres de gravité, leur appartenance à une classe définie devient beaucoup plus aléatoire jusqu'au point où il faut créer des zones grises étant donné qu'il est impossible de les rattacher à une classe particulière. De plus, en regroupant les signatures autour d'un nombre limité de noyaux, on risque de laisser tomber les classes moins distinctes au niveau statistique mais pouvant jouer un rôle important dans l'analyse spatiale des phénomènes. À cause de l'imprécision des signatures, il est en effet important de pouvoir déceler la transition d'un noyau à un autre et de permettre à l'interprète, soit de rattacher les pixels à un groupe ou autre en se basant sur le contexte spatial ou selon l'expérience, ou encore de pouvoir former une classe transitoire entre deux noyaux bien définis.

La classification supervisée se sert d'échantillons pour établir les caractéristiques des classes; ces échantillons peuvent venir de données de terrain ou de zones homogènes sur une carte. Or rien n'indique que la taxonomie désirée correspond à une discrimination spectrale, ce qui occasionne une certaine confusion entre les classes voulues et les classes obtenues et qui peut laisser un grand nombre de pixels non classés. Ce manque de correspondance peut être dû: soit à une discrimination inhabituelle apportée par des longueurs d'ondes non visibles, à un chevauchement de pixels sur des phénomènes différents, à un manque de discrimination entre les phénomènes à cause de signatures incomplètes, à la difficulté de trouver des zones homogènes, ou encore à la difficulté de trouver des échantillons représentatifs des classes (Goodenough 1974, Fontanel et coll. 1976, Braconne 1974).

LA MÉTHODE PROPOSÉE

La cartographie thématique rencontre certains problèmes en télédétection à cause de l'imprécision des données de base empêchant l'identification exacte des phénomènes géographiques et de la difficulté des méthodes d'analyse à extraire les différents thèmes des données spectrales. Certains de ces problèmes sont liés à la technologie actuelle et échappent au contrôle des géographes; cependant, c'est le rôle du géographe d'assurer que les méthodes utilisées dans l'analyse des données soient celles qui répondent le mieux aux objectifs de l'étude. Ce rôle est assuré d'abord par une connaissance précise des données de base, par une évaluation critique des méthodes d'analyse et enfin par le choix d'une sortie cartographique adaptée au type d'information présenté.

Ainsi, l'analyse d'un paysage spectral s'effectue en trois étapes: l'analyse des données de base afin d'évaluer la séparabilité spectrale des signatures, l'analyse du contexte géographique afin de saisir la correspondance entre la discrimination spectrale et la réalité du terrain et enfin la production de la carte thématique en tenant compte des taxonomies spectrale et géographique. Cette approche nécessite donc d'allier d'une part l'analyse quantitative à l'analyse qualitative, pour joindre l'expérience de l'interprète aux facilités de traitement informatique, et d'autre part d'allier la méthode de classification supervisée à la méthode non supervisée pour assurer une correspondance entre le monde spectral et le monde géographique. Pour illuster les différentes étapes de l'analyse on se servira de deux contextes géographiques différents: celui de l'île du Roi Guillaume, situé dans les Territoires du Nord-Ouest canadien et celui de Thetford Mines dans la région des Cantons de l'Est du Québec (voir figure 1). L'île du Roi Guillaume servira d'exemple de cartographie des foramtions géologiques de surface alors que la région de Thetford Mines servira d'exemple de cartographie de la végétation.

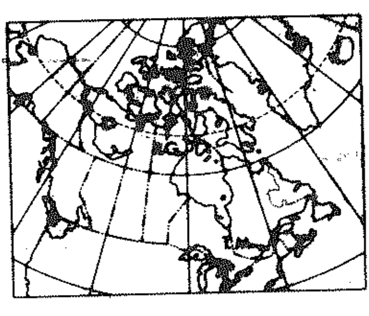


Figure 1: Localisation des régions de Thetford Mines et de l'île du Roi Guillaume

L'analyse des données de base

L'analyse des données de base à pour objet d'étudier le paysage au niveau spectral, c'est-à-dire d'étudier les relations qui existent entre les différentes composantes du paysage géographique et les signatures spectrales. Il s'agit d'une première étape de la classification des pixels en se basant principalement sur la discrimination spectrale. Pour ce faire, on considèrera les signatures comme des vecteurs à quatre dimensions dont chacun des axes correspond à un canal MSS. Pour faciliter l'étude des relations qui existent entre les signatures spectrales on se sert de corrélogrammes, réduisant ainsi à deux dimensions la représentation des vecteurs à quatre dimensions.

Les figures 2a et b présentent les corrélogrammes des différents canaux MSS des scènes de l'île du Roi Guillaume (Figure 2a) et de Thetford Mines (Figure 2b). Les figures 3 et 4 expliquent la relation entre les différentes données de terrain (taxons géographiques) et les nuages statistiques (taxons spectraux). À partir de ces corrélogrammes on peut observer les points suivants:

- Il n'est possible d'identifier précisément qu'un nombre très restreint de classes spectrales (la terre, l'eau et la glace sur la Figure 3, la terre et l'eau sur la Figure 4), les autres phénomènes étant englobés à l'intérieur de ces nuages.
- Il existe une forte corrélation entre les canaux 6 et 7 (proches infra rouge) sur toutes les scènes; les signatures sont cependant mieux réparties le long de l'axe du canal 7 que le long de l'axe du canal 6, permettant ainsi une meilleure séparation des signatures en se basant sur le canal 7. Cette corrélation permet d'éliminer le canal 6 lors de la classification des pixels, facilitant ainsi le traitement sans affecter grandement la classification.
- Les classes spectrales établies à l'aide de données de terrain ne correspondent pas nécessairement à des noyaux ou nuages spectraux, de plus la limite spectrale entre les phénomènes géographiques est imprécise et les signatures spectrales de phénomènes différents peuvent se chevaucher.

L'étude des corrélogrammes permet donc de tirer certaines conclusions en ce qui a trait à la classification des signatures:

- La taxonomie ne peut être basée uniquement sur l'analyse statistique des nuages spectraux mais doit être secondée par des données de terrain afin de voir la correspondance entre la discrimination spectrale et la taxonomie géographique.
- La limite entre les classes est souvent imprécise créant ainsi des zones transitoires entre les classes principales; l'impossibilité d'établir l'appartenance de certaines signatures à une classe définie oblige l'interprète à se référer au contexte géographique pour rattacher ces classes à un thème donné.
- Puisque les phénomènes géographiques ne constituent pas nécessairement des phénomènes discrets sur le plan spectral, il est donc important de reproduire les classes spectrales de façon analogique selon le niveau de réflectance et ainsi permettre de suivre la transition entre les classes.

L'analyse de la carte analogique

L'analyse des corrélogrammes a permis de saisir la relation texturale qui existe entre les différentes classes spectrales, l'analyse de la carte analogique a pour but d'étudier le paysae spectral à partir de la structure des classes. La structure des classes, rendue par une reproduction analogique des données spectrales, cherche à révéler les différentes formes du paysage et de permettre au géographe de faire appel à son expérience, en se servant du contexte régional, pour identifier certaines classes spectrales qui n'on pu être identifiées lors de l'étude des textures.

La reproduction analogique des images est obtenue en utilisant trois canaux MSS auxquels on a assigné les trois couleurs primaires (jaune, cyan et magenta) et en faisant correspondre l'intensité des couleurs au niveau de réflectance. On utilise généralement les canaux 4, 5 et 7 à cause de la forte corrélation entre les canaux 6 et 7. Pour assurer un maximum de contraste entre les différentes classes spectrales et faire correspondre les classes spectrales à des thèmes géographiques, on fait coincider les changements de teintes avec les limites des classes spectrales. En procédant ainsi, chaque classe spectrale correspond à une boîte à l'intérieur de l'hyperboîte formée par les trois canaux MSS; en termes informatiques chaque classe spectrale correspond à une variable d'une matrice à trois dimensions. Cette façon de procéder permet non seulement d'identifier sur l'image les classes définies à l'avance à l'aide de corrélogrammes, mais permet aussi de suivre l'évolution des classes transitoires entre les principales classes. Les classes transitoires peuvent être soit rattachées à des thèmes géographiques ou laissées tel quel si elles ne correspondent pas à un thème précis.

La carte thématique

La carte thématique regroupe les classes spectrales selon une taxonomie définie à l'avance. Le regroupement se fait à l'aide de l'image analogique sur laquelle on a identifié les classes se rattachant aux différents thèmes géographiques. Sur le plan technique, la carte thématique est produite en assignant des couleurs identiques à chaque point de la matrice ayant une signification géographique voisine; de cette façon le nombre de couleurs originales présentes sur l'image analogique est réduit au nombre de classes voulues. La réduction du nombre de classes permet d'assigner des couleurs plus contrastées aux différents thèmes choisis et de donner des couleurs intermédiaires aux classes transitoires.

L'étude de la carte thématique (exemple figure 5) permet d'illustrer certaines caractéristiques de la taxonomie spectrale et la façon qu'il faut interpréter les classes; en voici les grandes lignes:

- La discrimination spectrale fait ressortir certains aspects du paysage qui n'avaient pas été prévus au départ; cette télétaxonomie peut être parfois superflue mais en autres cas très bénéfique. Sur la carte de l'île du Roi Guillaume, la discrimination spectrale a fait ressortir plusieurs catégories de glace (vive, contenant des sédiments, fondante...) et d'étendues d'eau (claire et profonde, peu profonde, avec sédiments...) mais puisque ces classes ne présentent aucun intérêt sur le plan géologique on a dû les regrouper en quelques classes représentatives facilitant ainsi la lecture de la carte. D'autre part, la discrimination spectrale a fait ressortir sur la carte de Thetford Mines une classe spectrale non prévue mais qui s'est avérée très importante dans l'étude du paysage. Il s'agit d'une zone de végétation non identifiée recouvrant une superficie de plusieurs kilomètres carrés situés au sud-est de la ville de Thetford Mines. L'étude sur le terrain a révélé qu'il s'agissait d'une anomalie géochimique affectant la croissance de la végétation se traduisant par un faible taux de production de chlorophylle. Les méthodes de cartographie conventionnelles n'avaient pas décelé cette anomalie puisque le taux de chlorophylle produit par les plantes n'est détectable qu'à l'aide de l'infrarouge.
- Les classes spectrales transitoires entre les classes bien définies peuvent jouer un rôle important dans l'analyse du paysage. Les corrélogrammes de la scène de Thetford Mines ne révèlent pas la présence de trois classes de végétation. La subdivision de ce nuage a permis de cartographier les peuplements composés essentiellement de feuillus, de conifères, puis à un certain nombre de peuplements mixtes. Cette distinction est importante en cartographie forestière puisqu'elle permet d'évaluer le potentiel économique d'une forêt.

- Il est important de baser l'interprétation du paysage spectral autant sur la structure des phénomènes que sur la texture des classes spectrales. Les phénomènes de dimension inférieure à un pixel ne peuvent être identifiés spectralement de façon précise et peuvent être rattachés à leur classe d'appartenance en se servant du contexte géographique environnant. Sur la carte de l'île du Roi Guillaume, certains phénomènes géologiques tels que les eskers et les crêtes morainiques ont des dimensions irrégulières et leur signature spectrale peut varier en certains endroits; l'interprète peut cependant regrouper les pixels à l'intérieur d'une seule unité en se basant sur la structure du phénomène.
- Le choix des couleurs pour représenter les différentes classes sur la carte thématique joue un rôle important dans l'interprétation du paysage. L'étude des corrélogrammes a montré que les signatures spectrales ne forment que très rarement des classes discrètes et qu'elles ont plutôt un aspect continu, dont l'évolution se fait progressivement d'un noyau à l'autre. Il est donc important dans la reproduction des paysages spectraux de conserver l'aspect continu des classes spectrales afin de suivre l'évolution des classes. Sur la carte géologique par exemple, on représente la roche en place en blanc, les champs de felsenmeer ou de till délavé en jaune et les formations de till grossier en orange puisque ces formations géologiques ont des signatures spectrales très voisines et qu'il n'y a pas de limites précises entre ces trois classes. En donnant des couleurs contrastées à ces classes le géologue n'aurait pas pu suivre l'évolution des formations et l'interprétation des zones marginales aurait été très difficile à cause d'une alternance très rapide de couleurs contrastées.

CONCLUSION

Bien que très sommaire, l'étude des scènes de LANDSAT de Thetford Mines et de l'île du Roi Guillaume nous permet de dégager certains principes dont il faut tenir compte dans la cartographie thématique en télédétection:

- La cartographie thématique en télédétection diffère de la cartographie habituelle car la taxonomie doit répondre à une discrimination spectrale entre les phénomènes plutôt qu'à une discrimination décidée à l'avance par le cartographe.
- Contrairement aux cartes thématiques habituelles où chaque composante constitue une unité discrète, les classes spectrales forment les phénomènes continus dont la limite est souvent imprécise.
- L'interprétation des scènes spectrales se fait autant à partir de la structure des phénomènes que de la texture des classes spectrales, d'où la nécessité de reproduire de façon analogique les scènes spectrales.
- Il est nécessaire de comprendre la façon dont les classes spectrales sont formées et l'interprète doit intervenir dans la formation des classes; une classification aveugle basée uniquement sur une analyse mathématique peut donner des aberrations au niveau de la formation des classes.

Il est possible de produire des cartes thématiques à partir de données de télédétection mais l'interprète doit être conscient des limites de la discrimination spectrale et doit comprendre non seulement la nature des données de base mais aussi les méthodes de traitement de ces données. Contrairement à la carte thématique conventionnelle, les cartes thématiques en télédétection ne constituent que des documents intermédiaires nécessitant une interprétation plus poussée afin de répondre à une taxonomie voulue plutôt qu'à une taxonomie spectrale.

RÉFÉRENCES

Alfoldi, T. 1978, Introduction to Digital Images and Digital Analysis Techniques, Canadian Centre for Remote Sensing, Energy, Mines and Resources Canada, Technical note 78-1.

American Society of Photogrametry, 1963, Basic Matter and Energy relationships involved in remote reconnaissance, Photogrametric Engineering 29 (5), pp. 761-797.

Bélanger, J.R. 1978, Sur la synthèse d'image cartographique: Que signifie la juxtaposition de faibles et fortes fréquences de pixels dans les histogrammes de niveaux de réflectance?, Recherches Géographiques n° 8, ISSN 0396-9657, Strasbourg, France, pp. 75-86.

Braconne, S., Fontanel, A., Guy, M., et Lalemand, C., 1974, Cartographie thématique automatique: Principe et réalisation partielle d'un système; Proceedings Symposium on Remote Sensing and Photo-interpretation I.S.P., Banff, Canada.

Brun, F., Fontanel, A., Lalemand, A., Legendre, G., Rivereau, J.C., Thomas, G., 1974, Comparaison de différentes techniques de classement de données multispectrales, application à la géologie, Symposium EFRO, Frascati.

David, D.J., Joly, G., Verger, F., 1977, Le traitement par ordinateur des données de télédétection et leur cartographie automatique, 4e Symposium canadien sur la télédétection, Québec, Canada.

Diday, E., 1972, Optimisation en classification automatique et reconnaissance des formes, IRIA, note 6, n° 12.

Fontanel, A., Lalemand, C., Legendre, G., Rivereau, J.C., Thomas, G., 1976, Comparaison des images et les classifications multispectrales obtenues à partir de satellites LANDSAT, SKYLAB et du scanner aéroporté Daedalus, Journée de télédétection du JDTA, Toulouse, France.

Goodenough, D., 1976, Image 100 Classification Methods for ERTS Scanners data; Canadian Journal of Remote Sensing, v. 2, no. 1, Ottawa, Canada.

Goodenough, D., Shlien, S., 1974, Automatic Classification Methodology, C.C.R.S., Energy, Mines and Resources Canada, Technical note 74-1.

Hajic, E.J., Simonett, S., 1976, Comparison of Qualitative and Quantitative Image Analysis, Remote Sensing of Environment, Adison-Wesly Pub.

Lintz, J. Jr., Simonett, S., 1976, Remote Sensing of Environment Adison Wesley Pub., Mass., U.S.A.

Muehrcke, P., 1972, Thematic Cartography, Commission on College Geography, Resource Paper no. 19, Ass. of Am. Geog., Washington, D.C.

Reeves, R.J., Harris, G. Jr., 1977, Digital Processing to Aid Interpretation of Landsat Images, Journée de Télédétection du GDTA, St-Monde, France.

Rimbert, S., Langellé, J., 1969, Vers une automatisation de la cartographie thématique, Département de géographie, U. d'Ottawa.

Steiner, D., 1972, Automatic Processing and Classification of Remote Sensing Data, Geographical Data Handling, IGU Commission on Geographical Data Sensing and Processing, UNESCO/IFU Symposium, Ottawa, Canada.

Taylor, M.M., 1973, Principal Component Color Display of ERTS Imagery, Defence and Civil Institute of Environmental Medicine, Research Paper 73-RP-987A, Downsview, Canada.

Tomlinson, R.F., 1972, Geographical Data Handling, IGU Commission on Geographical Data Sensing and Processing UNESCO/IGU Symposium, Ottawa, Canada.

Turner, R.E., 1973, Atmospheric Effects, Remote Sensing of Earth Resources, v. 3, F. Shakrokhi (ed.).

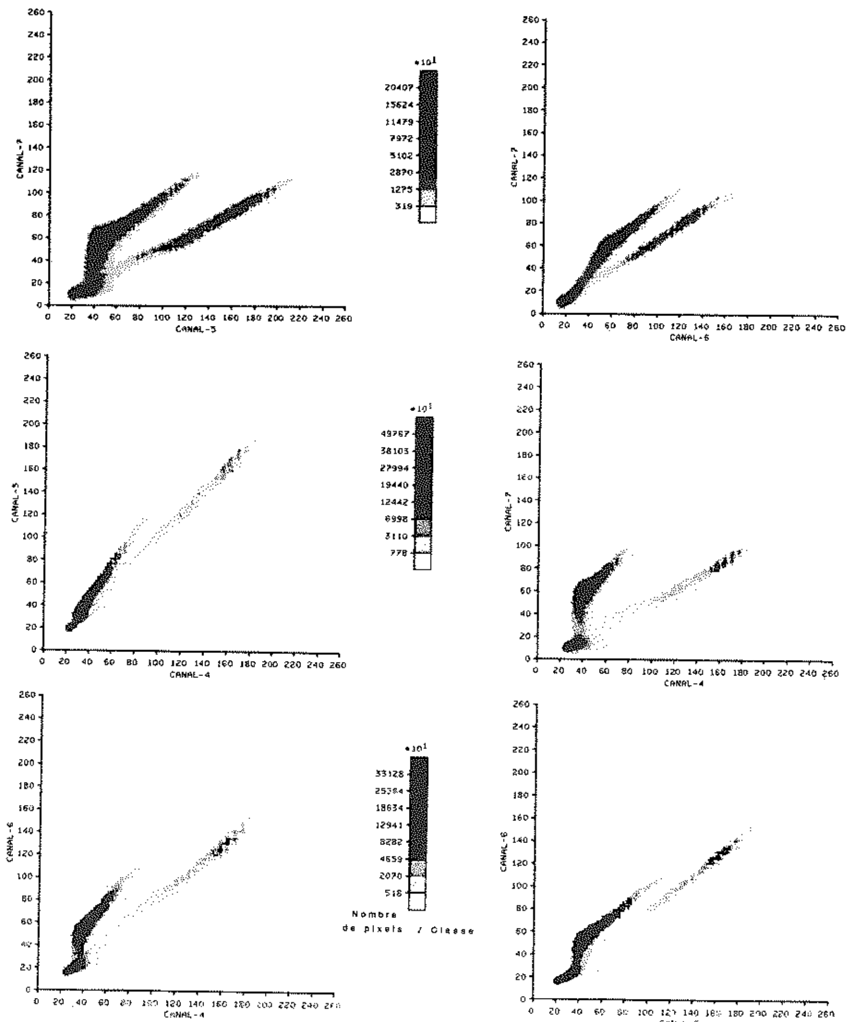
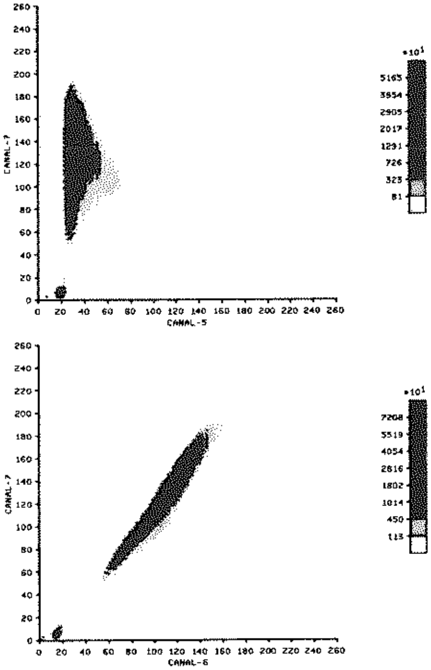
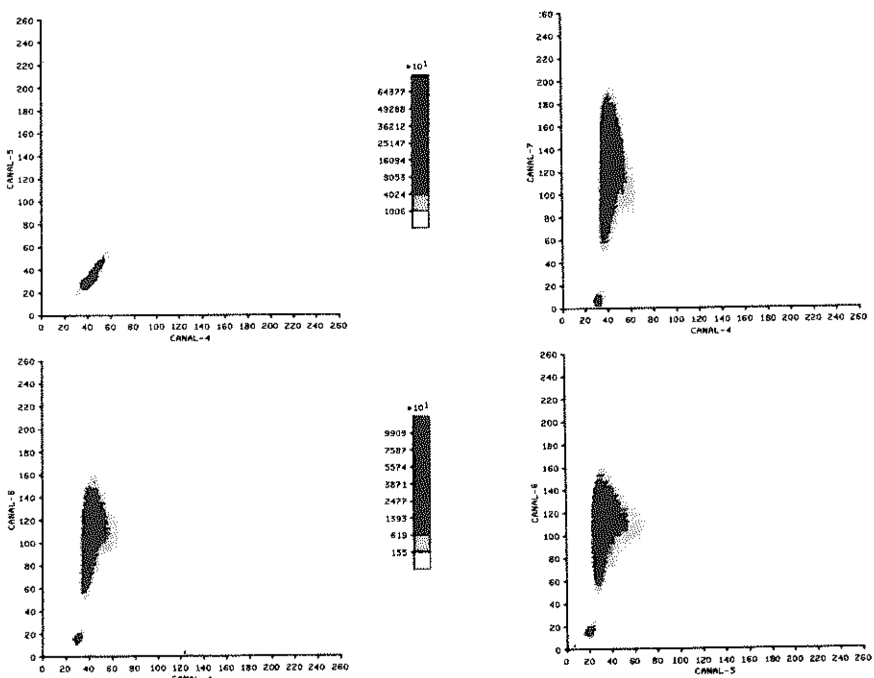


Figure 2a: Corrélogrammes des canaux MSS de la scène de l'ile du Roi Guillaume

Figure 2b: Corrélogrammes des canaux MSS de la scène de Thetford Mines





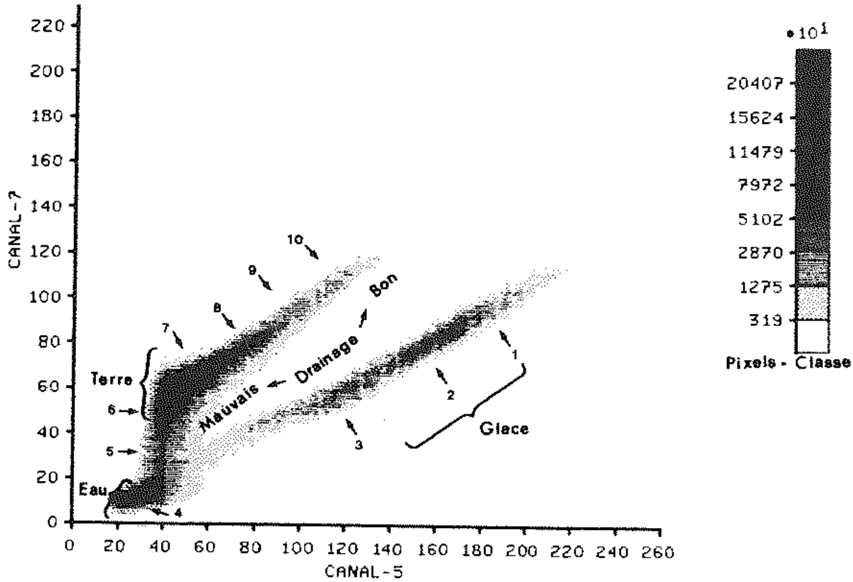


Figure 3: Identification des données de terrain sur les nuages spectraux des canaux 5 et 7, scène de l'île du Roi Guillaume.

1. Glace vive	6. Sédiments marins/végétation
2. Glace avec sédiments	7. Till bien drainé
3. Glace fondante	8. Till grossier, remanié
4. Eau profonde	9. Champ de blocs, galets
5. Eau peu profonde	10. Roche en place

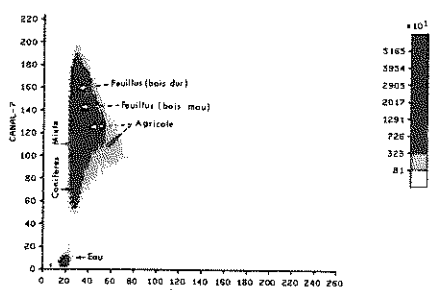


Figure 4: Identification des données de terrain sur les nuages spectraux des canaux 5 et 7, scène de Thetford Mines.

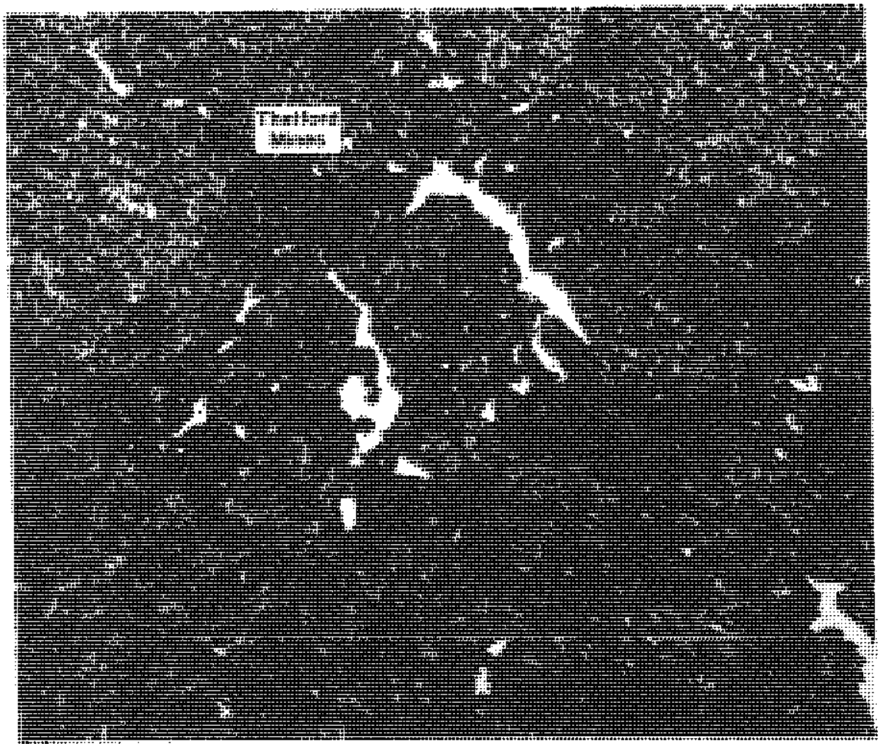


Figure 5: Carte de Thetford Mines. (Reproduction en noir et blanc d'une carte thématique en couleur).
Les peuplements de feuillus apparaissent en gris pâle, les conifères en noir et les peuplements mixtes en différentes teintes de gris.

A HIGH-ACCURACY AIRBORNE DIGITAL LINE IMAGER

R.A. Neville, S.M. Till, W.D. McColl, J.R. Gibson
Canada Centre for Remote Sensing
2464 Sheffield Road
Ottawa, Ontario
K1A 0Y7

ABSTRACT

An advanced airborne pushbroom scanner, MEIS II, has been developed and operated by the Canada Centre for Remote Sensing. The imaging scanner uses charge-coupled device linear array detectors, and covers the near infrared and visible spectral regions, spectral bands being readily selected by means of optical filters. MEIS II has a low noise equivalent radiance and high spatial resolution, and its real-time data processor provides radiometric gain and offset corrections, together with geometric and aircraft roll compensation. The sensor shows an improvement in noise equivalent radiance of at least two orders of magnitude when compared to the rotating-mirror scanner presently in use at CCRS as well as improvements in the geometric fidelity and spatial measurements, with inter-channel registration to within a fraction of a pixel.

INTRODUCTION

The Canada Centre for Remote Sensing is presently completing the first year of operation of a new solid state array imager developed and fabricated under contract to MacDonald, Dettwiler and Associates Ltd. This airborne imager, known as MEIS II, (multi-detector electro-optical imaging scanner) consists of the camera head, the image data resampler, and the data processor (Figure 1). The system is

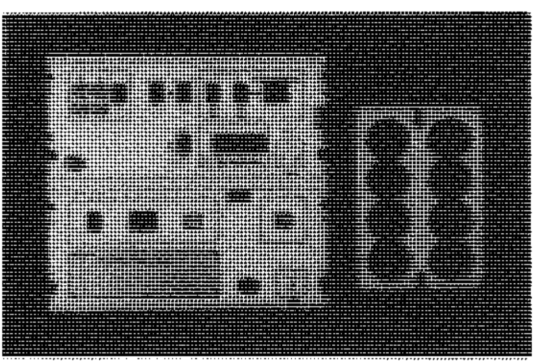


Figure 1: MEIS II Camera Head, Data Processor and Image Data Resampler

currently mounted in the Canada Centre for Remote Sensing Falcon 20 Fan-Jet, interfaces to a time code generator and an LTN-51 inertial navigation system, and outputs the image data, with navigation parameters merged, to a Bell and Howell high density digital tape recorder.

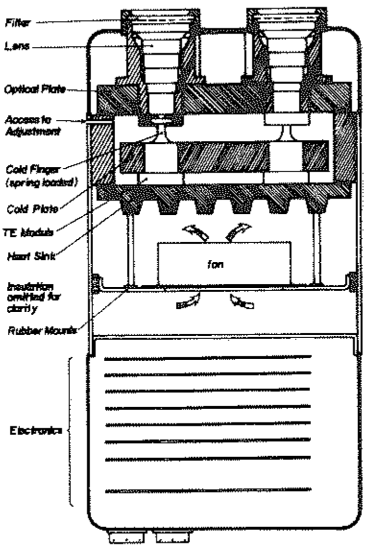
CAMERA HEAD

The camera head (Figure 2) has eight separate optical channels, each consisting of a 1728 element silicon charge-coupled device (CCD) linear array mounted in the focal plane of a high quality 35 mm format cine camera lens. The eight sets of optics and detectors are kept in precise alignment by an accurately machined optical bench face plate. The arrays are mechanically prealigned and fixed parallel to each other, boresighted on a distant target, and oriented such that in flight the linear dimension of each array is perpendicular to the direction of flight. Charges integrated in the individual detector elements are proportional to the light energy incident upon each photosite during the time between readouts. These charges are electronically sampled and converted to a series of digital signals which constitute an electronic line image of the scene below. The principle of operation of this type of imager is illustrated in Figure 3. The readouts of all photosites in all the arrays are performed simultaneously, thereby guaranteeing that the image line from all arrays are precisely superimposed and are individually free from motion induced distortion. The spectral bands detected by the sensor are determined by filters selected to match the application and mounted separately in front of each lens.

GEOMETRIC CHARACTERISTICS

While it is possible by mechanical adjustments to register the arrays to within 1/4 pixel in the flight direction it is not possible to do so in the across track direction because of the much larger size (1728 times) in this dimension. Channel-to-channel registration of pixels in the across track direction is accomplished by resampling the image data in real-time in the image data resampler (IDR). The IDR uses predetermined calibration data which effectively maps photosite locations in each array onto precisely determined look angles. In addition to providing the required registration to within 1/4 pixel this process also corrects for any lens distortions, and with the input of roll angle from an inertial navigation system, can compensate for aircraft roll. In the present configuration the IDR resamples from a portion of the 1728 detector pixels onto 1024 output pixels which are larger by the ratio of 4:3. The resampled field of view with a lens focal length of 24.6 mm is

39.66°, giving a pixel size of 0.70 mrad. In operation the appropriate aircraft altitude and ground speed are matched to the sensor image line rate, operator selectable from 25 Hz to 200 Hz, to give square pixels. MEIS II, mounted in the CCRS Falcon Fan-Jet, provides rectified imagery with pixel linear dimensions of 0.39 m to 7.4 m.



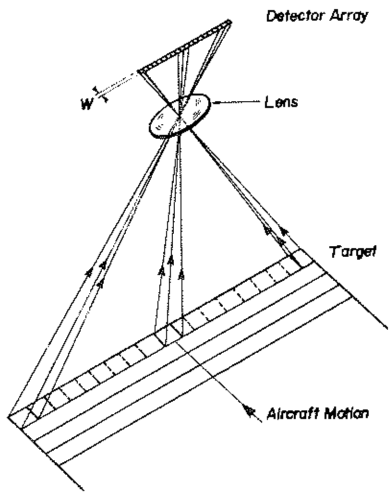


Figure 2: MEIS II Camera Head Vertical Section View

Figure 3: Principle of Operation of the MEIS II Imager

A stereo adapter is currently being developed which when attached to the sensor will deflect the field of view of one channel forward by 30°, another aft by 30°. A stereo image, either electronic or hard copy, is produced by superimposing the forward and aft images with the appropriate time offset. Should the need arise, this technique can be used to produce multispectral stereo imagery as well.

RADIOMETRIC CHARACTERISTICS

A radiometric calibration is performed for each filter set to correct for responsivity non-uniformities in each array and for the normal $\cos^4$ lens vignetting. The radiometric gain coefficients determined in this way and stored in the MEIS data processor memory are applied in real-time, along with dark current offset values measured in flight, to produce imagery that is free of the striping that would otherwise be present.

In addition to the greater geometric fidelity of the MEIS II imager as compared to moving mirror line scanners, MEIS II enjoys a much greater radiometric sensitivity over the whole silicon response range of 380-1100 nm. (Till et al., 1983). Its noise equivalent radiance, which is an inverse measure of sensitivity, is less than one hundredth that of a typical multispectral rotating mirror line scanner. The sensor dynamic range, defined as the ratio of the sensor saturation level to the RMS system noise, averages 9300:1 for an individual pixel and 6400:1 for a complete 1024 pixel long image line. This determines the range of illumination conditions over which acceptable imagery can be obtained for a fixed lens aperture setting. The intrascene dynamic range is determined by the digitization resolution; for MEIS II this range is 255:1.

APPLICATIONS

Operational missions flown to date have collected MEIS data for applications in cartography, water quality and water depth charting, forest classification, mapping of spruce budworm infestations, geological mapping, agricultural crop species identification and condition determination, and vegetation stress mapping, this last application requiring spectral bandwidths of only 3 nm.

Flights with MEIS in its fore-aft stereo mode are planned for the immediate future. The image data will be further processed post-flight to remove the effects of aircraft pitch and yaw, as well as roll. This procedure and subsequent operations to produce a digital terrain map are described in another presentation (Gibson et al., 1983) at this conference.

The advantages to cartography of a MEIS II type sensor are: i) the data is in digital form from sensor to final stage, eliminating the need for specialized equipment, ii) the high sensor sensitivity permits the acquisition of satisfactory data under low light conditions, iii) the sensor data has high geometric fidelity and is free of distortions introduced by the use of film, iv) the digital form of the image data is amenable to post-flight processing to correct the imagery for such problems as aircraft attitude induced distortions, variable intrascene illumination, and atmospheric path radiance.

REFERENCES

Gibson, J.R., O'Neil, R.A., Neville, R.A., Till, S. M., McColl, W.D, 1983, A Stereo Electro-Optical Line Imager for Automated Mapping, Sixth International Symposium on Automated Cartography, Ottawa, Ontario, 16-21 October 1983.

Till, S.M., McColl, W.D., Neville, R.A., 1983, Development, Field Performance and Evaluation of the MEIS II Multi-Detector Electro-Optical Imaging Scanner, Seventeenth International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan. May 9-13, 1983.

ANALYTICAL PLOTTER FOR FACILITY MANAGEMENT SYSTEMS DATA ACQUISITION

R. R. "Sky" Chamard
E. Coyote Enterprises Inc.
9688 Skye Drive
South Jordan, Utah 84065

Introduction

Modern facility management (FM) systems require data properly structured in a data bank for use in interactive displays, reports, and hard copy graphics. Acquiring data for these systems requires a very large amount of resources. In most FM systems the largest single cost element has been the data acquisition and the Automated Mapping (AM) phase. This phase, although expensive, is not always the most thoroughly thought out. As a result, the data in this basic structure package is of variable accuracies, comes from different data sources that are often not even tied to the same geodetic base, and are of different or unknown positional accuracy. FM systems can only be as good as the primary data in the AM system. The remaining of this presentation will dwell on base data (AM) requirements and effective methods of acquiring the data.

Automated Mapping

FM systems rely on an AM central unit of some kind. This usually contains all of the data needed to help in the facility management process.

Structuring of an AM system is critical to each particular user. The user must define:

1. Positional accuracy of the whole AM system.

2. Specifics of the type of data needed in the base, such as streets and roads, property lot location, power poles, manholes, curbs, location of underground facilities, trees, etc. etc.

3. Methods of information display, hard copy maps, orthophotos, interactive requirements, display by facility, composities, etc.

4. Coordinate display requirements.

5. Need for statistical reports and their formats.

6. Windowing requirements on displays.

7. Data update requirements.

8. Data security.

All of the above and often considerably more requirements must be addressed at the AM/FM project design phase.

When the specifications have been identified, a review of available facility positional data must be reviewed and verified. Map sheets and existing other drawings must all be evaluated for suitability of providing data to the required specifications.

Most utility companies have an existing manual mapping system to show where their installed facilities are located. This system can and usually consists of:

1. Design maps, often on linen; sometimes these are updated to show "as built".

2. Construction sketches of facilities.

3. Overall system maps, usually at a small scale.

4. Cadastral maps showing location of users, demands, and rate schedules.

5. Tax maps.

6. Others.

All of these maps and documents were usually developed to serve a specific single user demand, with very specific small area requirements. The eval-uation of the existing data is critical. It is easy to be pressured into utilizing existing data into the AM system. Lots of information exists and it should be utilized only if it doesn't degrade the requirements as set forth for in the AM specifications. Ignoring specifications leads to long-range catastrophic happenings in future decision making as applied to facility design and management.

Until the specifications for data requirements and the available data is thoroughly reviewed, it is too early to be convinced that some hardware system is required. It may be obvious at this state of the study that specifications are too rigid to be satisfied with existing maps, etc. Specifications must again be reviewed and, if still valid, a plan must be worked out to economically satisfy the AM data acquisition need.

AM systems covering large geographic areas have to be tailored differently than moderately-sized areas. A large system could be defined as one covering several states. A moderate system could be defined as a metropolitan area of less than 1,000,000 population. Both systems will require some of the same basic data and display systems -- methods of acquiring data for system input will vary.

System Overview - The data for system overview consists of a standard base that can cover the whole system in general terms. Items such as where is the location and power rating of all major electrical transmission lines, where are the location of major pipelines, main transportation routes and their relationship with the communities in the service areas, what types of land ownership do these traverse, political boundaries, etc.

In a large system, the base data for this AM base could be the U.S. Geological Survey's 1:500,000 scale state maps. Much of the data on these maps is available in a digital format. This data would provide the skeleton overview framework.

In a moderate or small size system, use of the U.S. Geological Survey's 1:24,000 quadrangle maps may be sufficient to provide the needed skeleton. It may be that the system is small enough that only one level of data base is needed. If that is the case, a totally new map or data base can be developed.

Both of the U.S. Geological Survey base maps and/or a new base can provide data sufficient in a skeleton form for the body of the complete AM to be augmented by other means.

System - "Detailed" Management Base - Moderate or small sized systems should not start their AM system with anything but a solid base. In most cases this will require an overall new mapping effort. Scale and accuracy requirements will have to be designed to meet the finest detail to be displayed. Mapping will have to be tied to a standard geodetic positioning system throughout. Most systems will suffice with a planimetric map with spot elevations of various features. A topographic map will not be necessary for the whole FM area, but may be specified for areas of particularly difficult terrain or areas in which complex engineering designs will be required in the future.

The system map must be structured in a layered format. This means one type of information is shown on one data layer. As an example, all above-ground powerlines on one layer, underground powerlines on a separate layer, streets and roads on one layer, cadastral data, etc. - each with their separate layer or residence in the computer system. Layers in a system can include:

- o Roads, streets, sidewalks;
- o Bridges and major highway structures;
- o Powerlines;
- o Ownership plats;
- o Sewerlines, waterlines, other pipelines;
- o Spot elevations;
- o Buildings;
- o Transformer locations;
- o Control monuments;
- o Hydrology - streams, ponds, etc.;
- o Political subdivision;
- o Others.

All layers tie together and crossing nodes can be identified by software for display purposes later in the AM process.

OATA ACQUISITION

Acquisition of data is necessarily a combined effort of gathering data visible on aerial photography and existing documents such as property descriptions, existing surveys of underground facilities, etc.

Existing Documents

Converting data on existing documents, such as maps, can be done by one of the following methods:

Hand Digitizing - Each map sheet is placed on a flat table that has "built in" an electrical impulse system - a digitizer. The operator follows each line with a curser and by so doing converts the line data to a series of X and Y coordinates. To do this task, the operator must key in the attributes of each feature.

An attribute for power lines, one for gas conduits, etc. This process is slow and labor intensive. It requires rigid edit of data prior to data entering into the AM system. Hand digitizing is the best method of digitizing existing maps on which all pertinent data has been shown by symbols or colors, but not by separate layers. This is the case in most instances of older existing documents.

Auto Scan - The auto scan process can be done with any of the many systems presently on the market. Scanning can be done by laser, raster, line following, or other systems. The auto scan process involves placing the document in a machine that scans the whole document in one pass or scans different layers of the document in separate passes. Scanning can be done in raster or vector modes. The scanned data is then computer processed into a format compatible to a particular AM format.

Both hand digitizing and auto scanning require considerable time to edit data to assure proper detail pickup and proper formatting for an AM system.

Creating New Data Base

The process of creating a new data base is an interesting process. It consists of the following steps:

Acquire Aerial Photography - Planning for aerial photography starts with the definition of what data is needed in the AM data base. The scale of aerial photography is based on being able to see the smallest detail needed to be mapped, map accuracy (positional accuracy), and the kind of photogrammetric instrumentation to be used. These factors are key to determining photography scale. Film emulsion is also a consideration to determine image definition - a decision will have to be made between the various black and white films and various color films.

Flight planning is done after the scale has been determined. Flight planning has to also consider factors such as fitting existing ground control and to best fit existing terrain, etc.

Ground and Office Control - In order to utilize aerial photography for map construction, it is necessary for known ground positions to be identified on the photographs and those positions tied to a geodetic base. To accomplish this task, existing known ground positions are identified on the photography and more identifiable positions are surveyed. If office procedures such as aerial triangulation are utilized, ground positions required can be held to a minimum. Office measuring methods are used to tie all of the aerial photography into a controlled block that can be used for mapping.

Mapping - Digitizing - Stereoplotters (instruments that can reconstruct the geometry of a photograph and record features on those photographs in proper positions) are used to produce maps. In recent years, these instruments have been modified to include encoders, black boxes, etc., to allow them to be used as three-axis digitizers. The procedure is quite simple. Adjoining photographs in which 60 to 65 percent of their coverage overlaps is set up in special photograph holders on the instrument. These are adjusted so the instrument operator can see the overlap area in three dimensions. Known positions are then keyed into the system and the operator can record all positions on the photographs as X, Y, and Z coordinates in the chosen geodetic control system. These positions can be "played back" on a screen (CRT) so the operator can see if everything on the photographs has been properly recorded. This process acquires data required for an AM system directly in proper format without having to draw or compile a map for later digitizing as previously discussed. In recent years, the analytical stereoplotter (AP) has been developed making this process more efficient and accurate.

The analytical stereoplotter is a stereoplotter that is designed as a total computer assisted instrument. Every instrument motion is a computer assisted coordinate generator. It is easier and more efficient to operate in a digital mode than conventional stereoplotters and is more accurate, as all factors involving geometry of the aerial photographs are mathematically determined. A typical analytical plotter is the US-2 as shown in Figure 1. Analytical plotters consist of the following components:

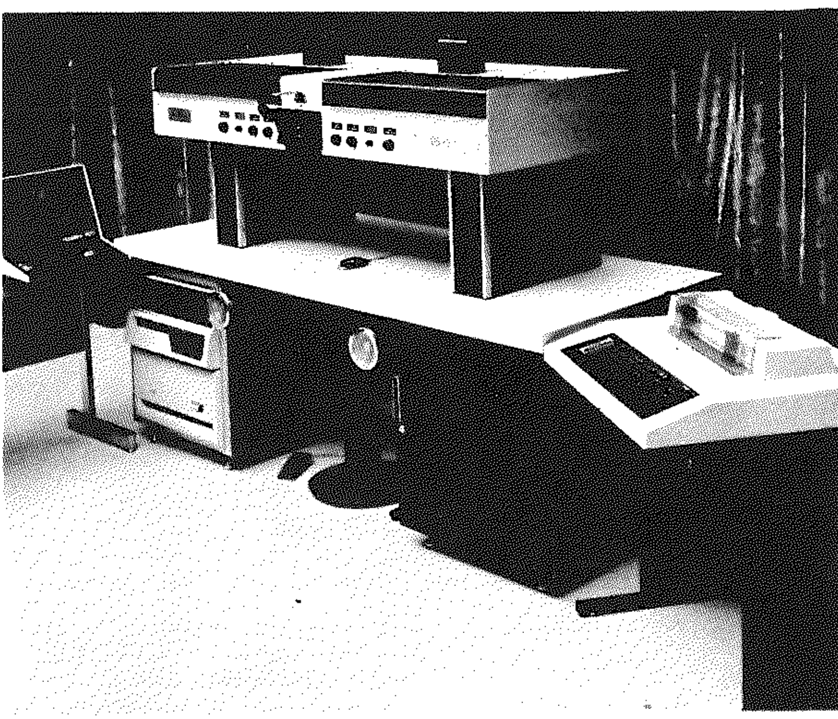


Figure 1

- A precision stereo viewer on which two overlapping aerial photographs are placed to form a stereomodel. This pair of photographs allows for three dimension viewing and for accurate measurements to be made. Measurements are made by recording X and Y coordinates of the same image on photographs 1 and 2. These measurements are made through the use of precise linear encoders and the photogrammetric transformation computations are made quickly by a high speed micro processor, in real time, converting those measurements to ground positions in any specified coordinate system.

- A high speed micro processor is designed to compute in real time all of the operations required to compute photogrammetric computations. In efficient analytical plotters, this micro processor is used to relieve the load on the host computer.

- Host computer - this is a mini computer of high quality and capability. It works in concert with the micro processor. It also provides all the application programs needed to make an analytical plotter efficient. This host is the medium for data transfer from the analytical plotter to the AM system.

- Operator station consists of a keyboard and CRT. This is the communication device on which the operator can command functions needed to be performed by the viewer, micro and host computer.

- Printer output - a teleprinting device which operates at high speed. It provides hard copy verification of the accuracy of stereo model setup and is useful to document operations. This teleprinting device is also an input station into the analytical system.

- Software - the analytical plotter's strengths can be evaluated by the array of application programs that are available. These programs provide the versatility of data transfer and application to any AM system.

COST OF ACQUIRING DATA WITH AP

Using the AP as the data gatherer, the following is typical of cost and times to acquire data:

AM/FM system requirements are, for example:

- Capability to make maps at 1" = 200' scale;

- Fire hydrants smallest feature to be seen and mapped;

- Typical area - the area to be mapped consists of a typical urban sprawl area with homes on .1 to .2 acre lots; streets, curbs, and sidewalks; facilities such as power poles, etc., above ground;

- Scale of aerial photography to meet requirements is 1:24,000;

- Digitizing and resulting maps to meet National Mapping Accuracy Standards.

Typical AP Production Times to Acquire Data

These times are only for data acquisition from photographs. Costs for acquiring aerial photography and field control would have to be added for estimating purposes.

Function	Time Per Model Hr.
Aerial Triangulation - Process consists of measuring control points and ties each photograph to adjoining ones and to a geodetic base - includes computer adjustments.	0.30
Model Reset - Place aerial photos in AP for digitizing	0.05
Planimetric Digitizing - Digitize all features such as power lines, houses, etc.	13.25
Transport Facilities - Roads, curbs, etc.	2.40
Hydrographic Details	1.00
Spot Elevations	0.75
Edit	5.00
Total hrs./model	22.75 hrs.
Total sq. mile model 3.25	
Time/sq. mile	7.00 hrs.
Average cost for AP w/operator $40/hr.	
Cost per sq. mile	$280.00
Cost per acre	$ .4375

Typical total cost AM data base acquisition for area of 1,000 square miles:

1. Aerial Photography	$ 12,500
2. Field Control	12,000
3. Data Acquisition w/AP	280,000
4. Software Systems	310,000
5. Final System Edit	195,000
Total Cost	$809,500
or	$ 8,095 sq. mile

These costs do not include cost of "in house" hardware such as a computer system, CRT stations, disk storage devices, and hard copy output systems. Depending on the requirements, these hardware systems are expensive.

Conclusion

AM system data acquisition is one of the most expensive components of an AM FM system. It is important that the process be done accurately for future use. Having data in a computer-based system can give a false feeling of accuracy that may not exist. This can lead to poor decision making in the future.

Basic mapping costs for the AM system can be effectively done by using the analytical plotter as the data gatherer. Cost shown in this paper of $.4375 per acre is very competitive versus other more traditional methods.

UNDERWATER MAPPING TECHNIQUES USING REMOTE SENSING FOR SALVAGING SUNKEN VESSELS

Dr. A. A. Navarro
NASA-NRC, PHILIPPINES INC.
Manila, Philippines 3119

ABSTRACT

If we have decided to use Photogrammetric Bathymetry in the exploitation of submerged lands' natural resources and concern with environment , wetland areas, littoral zones, and flood plains that have altered traditional mapping priorities and generated the development of new techniques to serve these new interests. We are in the right track for its used in sunken vessels salvaging. In the modern color aerial - photography , with its remarkable clearwater penetration -characteristics and dramatic presentation of submerged details, provides a basic tool and is a supplement for mapping the seabed in shoals and waters of moderate depths of the sea. Although photometric and spectral photograph density and co -lor measuring techniques show promise for mapping the general configuration of sandy beach slopes , at present the vertical accuracy requirements for nautical cherting can be assured through the use of precised three dimensional photogrammetric techniques. Although single stereoscopic models can occasionally be oriented for the compilation of depth curves and soundings around small islands of inlets, extensive offshore shoal areas require blocks of overlapping strips of aerial photographs to bridge the zones between the vertical control points.

INTRODUCTION

The theory and application of underwater mapping to sunken vessels is a solution of the two-media photogrammetric problem requires several minor departures from the method normally used when the bundle of light rays passes solely through the atmosphere . The effect of refraction at the water-air interface must be taken into account when aerotriangulation is to be performed where a portion or all of the imaged points are underwater. The solution of this problem requires the formulation of the mathematical model for two-media refraction (See Figure I)

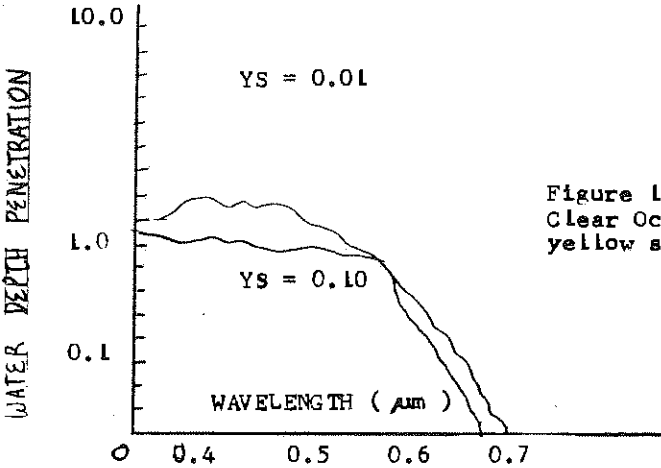


Figure 1 :
Clear Ocean with yellow substance

THE MATHEMATICAL MODEL FOR UNDERWATER MAPPING FOR SALVAGING SUNKEN VESSELS :

The basic mathematical model for treating the photograph image coordinates for the effect of two-media refraction on underwater points consists of the expression of the radial displacement which is the following : (Figure 2)

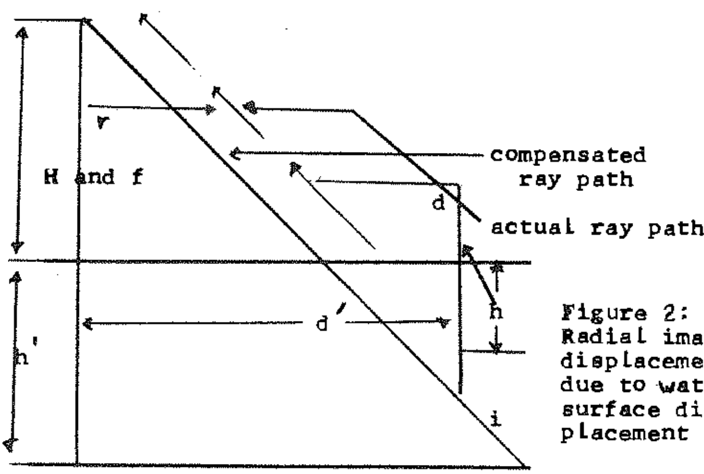


Figure 2:
Radial image displacement due to water surface displacement

THE MATHEMATICAL MODEL FOR UNDERWATER MAPPING:

Figure 2 :, and can, for vertical photographs, be expressed as :

$$\Delta d = \frac{dh\left(1 - \frac{1}{a}\right)}{H - h} \quad \text{(EQUATION 1)}$$

Where:

Δ d = is the required correction, in meters, to be applied to the photograph image. Its sign is always negative or toward the photocenter.

d = is the radius of the photograph image , in meters and is equal to $(X^2 + y^2)^{\frac{1}{2}}$

h = is the depth of the underwater point with respect to the surface datum at the time of photography. It is expressed in meters and is always negative in sign.

H = is the flying height ,in meters , with respect to mean sea level.

a = is the ratio of tangents of the angles of refraction (r) and incidence (i) (Rinner, 1969). It is dependent upon (d),(f), and (n)

(EQUATION 1)

$$\Delta d = \frac{dh\left(1 - \frac{1}{a}\right)}{H - h}$$

$$\Delta d = \frac{56 \text{ KMS.} \times 200 \text{ KMS.} \left(1 - \frac{1}{25{,}000 \text{ KMS.}}\right)}{1{,}044 \text{ kms.} - 200 \text{ KMS.}}$$

$$\Delta d = 5\ \mu m$$

USE OF REMOTE SENSING IN WATER RESOURCE MANAGEMENT *

Remote Sensing techniques can be applied to water resource decisions at many stages of development . However, successful utilization of these techniques involves a number of decisions ranging from the purely technical evaluation of sensor capabilities to economic trade-offs and organizational adjustments.

Typical Applications of Remote Sensing

It will describe typical applications of remote sensing to planning water resource developments, designing a major water resource project, and operating and maintaining regional water systems . Emphasis is placed on management decisions needed at each step in the process to assure effective use of the technology.

Illustration No.I: Appropriate Technology for Salvaging.

*, Originally published in Twelfth International Symposium on Remote Sensing of Environment; Manila, Philippines Jarman, J. W., 1978. This presentation by permission.

OPTICAL REMOTE SENSING OF CHLOROPHYLL IN OCEAN WATERS *

Ocean color remote sensing experiments have been conducted in coastal waters utilizing measurements from high altitude aircraft coordinated with supporting measurements from surface vessels . The results of these experiments have been used to study the significance of the various factors contributing to the apparent signals available to the remote sensor , to study mdthods for eliminating the masking effects of surface glitter and the atmosphere , and to develop effective methods for equating the remotely sensed signal with chlorophyll concentrations in the water.

Right: Schematic representation of the acoustic detection methods of a frigate.

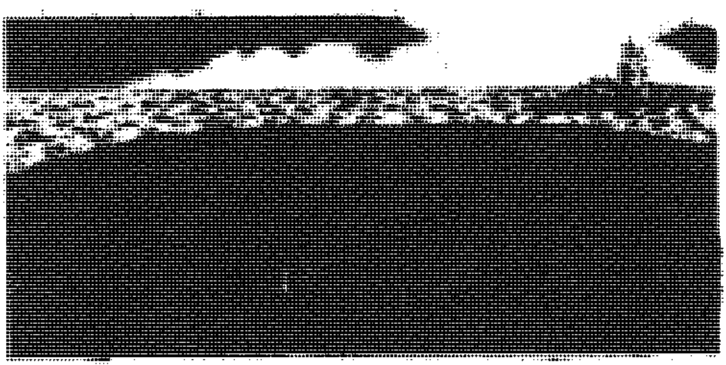


Illustration No. 2 : Underwater Mapping and Positioning of Sunken Vessels beneath Ocean.

---

* Originally Published in Remote Sensing of Environment Tweltfh International Symposium; Manila, Philippines , Austin, R.W., Wilson , W.H., and Smith , R.C. , 1978 . This presentation by permission.

IMAGE NOT PRINTABLE

ILLUSTRATION No.3 : REMOTE SENSING TECHNIQUES FOR UNDERWATER MAPPING AND LOCATING, SALVAGING SUNKEN VESSELS .

IMAGE NOT PRINTABLE

ILLUSTRATION NO.4 : UNIVERSAL APPLICATIONS OF REMOTE SENSING FOR LOCATING, MAPPING, AND SALVAGING SUNKEN VESSELS.

REMOTE SENSING OF THE OCEANS- A REVIEW FOR SUNKEN VESSELS SALVAGING *

By the use of active microwave sensors SEASAT-A is able to detect information at the ocean surface through cloud cover day and night albeit over a limited swath, between approximately 72°N and 72°S latitude.

Data Products will be in Significant Wave Height

In general data products will be significant wave height (sea state) , sea surface topography and geoid (from the short pulse radar altimeter -ALT); wind speed and direction (from the wind field scatterometer -SCAT); surface wave and current patterns, ice fields and coastal/ocean interaction (from the Synthetic Aperture Radar -S A R) ; Ocean Surface features recognition and surface temperature , atmospheric water vapour and ice coverage (from the Scanning Multichannel Microwave Radiometer (S M M R) .

SEASAT - A will have Three Microwave Radars

Thus, Seasat- A will have three microwave radars (ALT, SAR and SCAT) one microwave radiometer (SMMR) and one visual infrared radiometer (V I R R). Once SEASAT -A has demonstrated the potentials of a global ocean monitoring system , follow -up on operational series comprising multiple spacecraft will eventuate. These maybe in the form of low cost modular spacecraft (L C M S) presently being evolved for compatibility with the U.S. Space Transportation System - Space Shuttle.

---

* Originally Published in Twelfth International Symposium in Remote Sensing of Environment ; Manila, Philippines. Morgan , G.A., 1978. This presentation by permission.

OCEANWAVE IMAGERY AND WAVE SPECTRA DISTORTIONS BY SYNTHETIC APERTURE RADAR*

If a moving object is imaged by a synthetic aperture radar (S A R) is subject to focus errors, and even disappearance from the image. Whereas there are many examples of S A R oceanwave imagery for this imaging process, existing models do not adequately explain phenomena observed in radar images . The general purpose of this paper is to review and advance the theory of moving object S A R imagery, as it pertains to the sea surface, and to compare theoretical predictions with measurable observations based on actual S A R data.

Occurance of Distortions in the Apparent Direction

The most obvious effect is the occurance of distortions in the apparent direction and spatial dimensions of the scene elements in motion, a phenomenom that results when any scanning remote sensing system (e.g. multi-spectral scanner, laser line scanner or side- looking radar), is used to create an image of uniformly moving surface phenomena such as gravity wave swells. This distortions occur because the along-track scanning of a sensor either expands or contracts the along-track scale factor of an observe wave field, depending on whether the prevailing wave motion has a component of velocity in some direction as or opposed to the sensor velocity . The distortion becomes more severe as the ratio of wave velocity to sensor velocity increases, and is a function of wave direction relative to sensor direction . Principal distortions impact apparent wave direction and periodicity , and maybe substantial for oceanic swells observed from typical airborne remote sensing platforms . The paper quantifies the theory and comments on its implications.

---

* Originally Published in Tweltfh International Symposium on Remote Sensing of Environment ; Manila, Philippines , Raney , R.K. and Lowry, R.T. , 1978 . This presentation by permission .

CONCLUSION

Since the change of value, can occur and its effect on underwater vertical control points by 1°or 2°tilt in the aerial photograph is insignificant when the flying height is more than 100 times as great as the water depth , the refraction compensated coordinates of all vertical control points can be determined prior to analytic aerotriangulation using (Equation I) . This should be done following coordinate refinement for file and camera distortion and for comparator error.

ACKNOWLEDGEMENTS

The author is highly indebted to the American Society of Photogrammetry and to the Auto-Carto VI for granting this oppurtunity in presenting its paper and special thanks is due its staff and members of this two organizations.

REFERENCES

Cook, J.J.; 1978, SUMMARIES ; Tweltfh International Symposium on Remote Sensing of Environment , VOL. I-pp. 69-133

Deutsch,M.; 1980, LANDSAT DETECTION OF OIL FROM NATURAL SEEPS Photogrammetric Engineering and Remote Sensing, VOL. 46, pp. 1313 -1322

Mcnally,R. and Beazley, M. ; 1979, the nature of resources, MAPPING TECHNIQUES: OUR MAGNIFICENT EARTH , pp. 22 and 23 Published in the United States of America by Rand McNally & Company P.O. Box 7600, Chicago,Illinois 60680

Reeves, R.G.; Anson,A.;Landen,D.; 1974, The Marine Environment ; Manual of Remote Sensing, VOL. II -pp. 1574- 1577

COMPUTER-ASSISTED PHOTO-INTERPRETATION

Carl N. Reed
Autometric Incorporated
Fort Collins, Colorado 80526 U.S.A.

ABSTRACT

During the last two years, a new technology that marries computer technology with traditional photo-interpretation has been evolving. Still in its infancy, this evolution promises to have a considerable impact, especially in large mapping agencies, on how mappable information is collected, processed, stored, and displayed. Called CAPI, short for Computer Assisted Photo-Interpretation, digital processing, analytical photogrammetry, artificial intelligence, refresh graphics, voice interaction, and image processing hardware/software are being merged. The goal of this merger is to have the computer take the first cut at interpreting a stereo pair and then have the trained interpretor interactively rectify/correct the results of the computer's effort.

This paper will focus on the current state of this technology. In addition, projects using the first generation of CAPI will be discussed to show how the computer-interpretor synergism can be extremely effective. The paper will conclude with a discussion of current and future research on CAPI.

AUTOMATIC CARTOGRAPHY OF AGRICULTURAL ZONE BY MEANS OF MULTITEMPORAL SEGMENTATION OF REMOTE SENSING IMAGES

Eric CALS, Robert JEANSOULIN, J. Claude DARCOS
Laboratoire "Langages et Systemes Informatiques"
2 rue Camichel - 31071 TOULOUSE CEDEX - FRANCE

ABSTRACT

The need of new concepts and techniques to use the time dimension is growing rapidly as a consequence of the increasing number of multitemporal and multi-source images in remote sensing; the forthcoming French satellite SPOT will speed up the flow of high-resolution and repetitive data.

This paper focuses on the multitemporal segmentation and analysis of remote sensing images as a part of geometric reasoning and scene understanding, and on the utilization of these segmentation results for new classification schemes.

In the context of an agricultural experiment, the "Lauragais project", a simulation of visible and near-infrared spectral bands of SPOT HRV and Landsat-4 TM has been achieved from an aircrafted Daedalus Multi Spectral Scanner.

The following features are described:

- individualization of entities (parcels of land) on each mono-temporal image by multispectral segmentation based on fuzzy sets approach,

- cartography of three test-sites: we plot the contours of each segmented entity into the appropriate geometric representation. These geometric descriptions are compared from date to date to give a multitemporal description of the landscape.

- automatic training for a pixel-by-pixel supervised classification using a gaussian maximum likelihood criterion for crop inventory. A comparison between the results of the classifications with this automatic method and a manual method for training has been performed.

- entity-by-entity supervised classification with a dynamic cluster method using non-parametric characteristics of probability density functions of each entity.

LIMITATIONS OF LANDSAT IMAGERY FOR THEMATIC URBAN MAPPING

Mohammad Id Ozone
King Saud University, College of Engineering
P.O. Box 800, Riyadh 11421
Saudi Arabia

ABSTRACT

In a research project, at Kuwait Institute for Scientific Research, for monitoring the expansion of Kuwait city during the seventies, thematic urban maps are compiled on overlays scale 1:100000 from a LANDSAT scene. Themes of interest are "thoroughfares, urban areas, uninhabited areas and land cover". Compiled features are field identified as well as compared to maps valid at the date of the scene, and found satisfactory.

The overlays are also superimposed, and an integrated city map is reproduced to the same scale 1:100000.

Thematic maps 1:60000 and larger are endeavored. Such relatively large scales are found not to meet the requirements of the project, due to poor resolution (80 m) of LANDSAT 1 imagery available for that period. Both scales are compared indicating deficiencies and anomalies on the larger scales.

Analysis of the deficiencies and anomalies appearing on enlargements of old generation LANDSAT scenes shows that they can be eliminated on the new generation LANDSAT 4, having 30 m resolution thematic mapper with the addition of the thermal band and should meet most of the requirements needed for efficient monitoring of the expansion of the city.

ELECTRONIC EXCHANGE/DISTRIBUTION OF CARTOGRAPHIC DATA

An Intermediate-Scale Digital Cartographic Data Base for National Needs 563
Stephen Guptill

Computer-Assisted Mapping for the Census of Canada 570
D. Ross Bradley

The Availability and Use of Digital Topographic Data 580
M. Rodrigue and L. Thompson

Two Way Data Transfer Between Aries Image Analysis System and ARC-INFO Geographic Information System 588
Robert Maher, David Colville and Douglas Rigby

On the Transfer of Remote Sensing Classifications Into Polygon Geocoded Data Bases in Canada 598
D.G. Goodenough, G.W. Plunkett and J.J. Palimaka

A Demonstration Transfer of Remotely Sensed Data Utilizing the Standard Format for the Transfer of Geocoded Polygon Data 607
Kenneth Korporal

Cartographic Feature Coding 616
Bruce Wood and David Douglas

The Archiving of Computer Cartography (Abstract) 627
Louis Cardinal and Betty Kidd

LATE ARRIVAL(S)

IGDMS: An Integrated Geographic Data Management System (Abstract) 628
Timothy Nyerges

AN INTERMEDIATE-SCALE DIGITAL CARTOGRAPHIC DATA BASE FOR NATIONAL NEEDS

By Stephen C. Guptill
U.S. Geological Survey
521 National Center
Reston, Virginia 22092

ABSTRACT

The U.S. Geological Survey is beginning a major new program to create a nationwide digital cartographic data base from 1:100,000-scale maps by the end of the decade. This data base will supplement the currently available 1:2,000,000-scale national data base and the selected coverage at 1:24,000-scale. It will provide complete coverage of transportation features (roads, railroads, powerlines, pipelines) and hydrographic features (streams, rivers, water bodies). It is anticipated that the data will be useful for both the production of custom graphics and as basic input to geographic information systems.

The decision to create this data base was reached, in part, through the combination of two forces: the maturing of efficient and economical data capture technology and the requirements expressed for a major use of the data. The 1:100,000-scale map series was designed to facilitate automated data capture (for example, by raster scanners). The methodologies and procedures used in the data capture process have developed enough so that we can take advantage of the design features of the 1:100,000-scale maps and rapidly build a digital cartographic data base. Once these data are produced, the U.S. Bureau of the Census intends to use them as the cartographic framework of their geographic support system for the 1990 Decennial Census. The collaboration of a major data producer with a major data user at the inception of this project augurs well for the successful creation of an intermediate-scale digital cartographic data base that will meet national needs.

INTRODUCTION

For a number of years the U.S. Geological Survey (USGS) has been pursuing the development of a Digital Cartographic Data Base (DCDB). The initial plans were for the data base to consist of boundaries, public land net, streams and water bodies, and transportation features shown on 1:24,000-scale maps; elevation data largely obtained concurrently with the orthophotoquad program; planimetric features from the 1:2,000,000-scale sectional maps of the National Atlas of

Any use of trade names and trademarks in this publication is for identification purposes only and does not constitute endorsement by the U.S. Geological Survey.

the United States of America; elevation data obtained from the 1:250,000-scale map series; land use and land cover and associated map data; and geographic names. Because of a number of technological developments and programmatic opportunities, the scope of the data base has been increased to include digital cartographic data collected from 1:100,000-scale base maps. These data will provide complete coverage of transportation features (roads, railroads, powerlines, pipelines) and hydrographic features (streams, rivers, water bodies). This major new effort will attempt to achieve nationwide coverage by the end of the decade.

CONTENT

Data Base Description

1:100,000-Scale Maps. In the mid-1970's the USGS began its program of 1:100,000-scale mapping. The level of content was determined mainly by evaluating the content of the 1:24,000-scale maps. Each feature shown at 1:24,000 was evaluated and designated "include" or "exclude" at 1:100,000. The drawing (feature separate) on which the feature would be shown was also determined. The color-separation system of map production, in which a separate drawing is prepared for each color to be printed, was expanded to a feature-separation system. In the new system the major components of the map (for example, roads and hydrography) are subdivided into classes, and separate master scribed drawings are prepared for each class.

The framework and content of the 1:100,000-scale base maps is derived primarily from 1:24,000-scale maps, with updates during the production process. This production process has five phases: (1) reducing larger scale maps to 1:100,000-scale and mosaicking on the UTM projection; (2) updating the mosaicked base; (3) scribing the planimetric feature separate manuscripts; (4) scribing the contour manuscript; and (5) printing the complete metric topographic edition. If the base map material is less than 3 years old, only major features are updated. Material older than 3 years often requires more extensive planimetric updating. High-altitude aerial photographs are frequently used as the source of the update.

Features Included in the 1:100,000-Scale DCDB. The data base will provide comprehensive coverage of transportation and hydrographic features. A complete list of the features to be included is given in table 1. These features provide a basic framework of geographic information needed by major users (such as the U.S. Bureau of the Census) to perform various types of spatial analyses.

Data Structure

The digital planimetric data are produced and distributed in the form of digital line graphs (DLG's). The DLG concept is based on graph theory in which a graph can be represented as a set of nodes and links that explicitly records the spatial relationships inherent in the graph.

Table 1.--Features included in 1:100,000-scale digital data base

Transportation Features	Hydrographic Features
Airport, airfield, landing strip	Alkali flat
Heliport	Aqueduct, conduit, flume (elevated)
Pipeline	Aqueduct, flume, etc. (underground)
Power substation	Aqueduct tunnel
Power transmission line	Area subject to controlled inundation
Railroads and related features	Area to be submerged
Bridge	Breakwater, pier, wharf
Carline	Canal, flume, aqueduct, or perennial ditch
Ferry	Canal, intermittent
Roundhouse	Canal lock or sluice gate
Sidings	Canal, navigable
Snowshed	Channel in water area
Station	Cranberry bog
Tracks, narrow gage	Dam masonry
Juxtaposition	Dam with lock
Multiple	Ditch intermittent
Multiple, abandoned	Drydock
Multiple, dismantled	Dry lake or pond
Multiple, under construction	Falls
Single	Filtration plant
Single, abandoned	Fish hatchery
Tracks, standard gage	Gaging station
Juxtaposition	Glacial crevasses
Multiple	Glacial or permanent snowfield
Multiple, abandoned	Lake or pond, intermittent
Multiple, dismantled	Lake or pond, perennial
Multiple, under construction	Lock, shipping canal
Single	Mangrove
Single, abandoned	Marsh or swamp
Single, dismantled	Rapids
Single, under construction	Reservoir
Tunnel	Salt evaporator
Underpass, overpass	Seawall
Yards	Sewage disposal
Roads and related features	Shoreline
Bridge	Siphon
Class 1	Spring
Class 2	Stream, braided
Class 3	Stream, disappearing
Class 4	Stream, intermittent
Class 5 (trail)	Stream, perennial
Dead-end road	Stream, unsurveyed perennial
Ferry	Submerged marsh or swamp
Interchange	Wash
Parking area	Watermill
Paved service and rest areas	Water surface elevation
Tunnel	Water well
Under construction, class 1	Windmill
Underpass, overpass	
Ski lift, tramway, incline railway	

The DLG topologically structured data file accommodates all of the categories of data, that is, point, line, and area data types, that are included in table 1. Each distinct major data category (such as transportation and hydrography) is stored as a separate data file in the data base. Details on the DLG structure can be found in USGS Circular 895-C (Allder and Elassal, 1983); listings and explanations of the attribute (feature) codes are contained in USGS Circular 895-G (Allder and others, 1983).

This data base, when complete, will contain an immense amount of spatial information. Not only are a large number of features identified and coded, but also the level of feature portrayal is quite detailed. Figure 1, a plot of a portion of the road network digitized from the Tallahassee, Florida, 1:100,000-scale map, provides an indication of this detail.

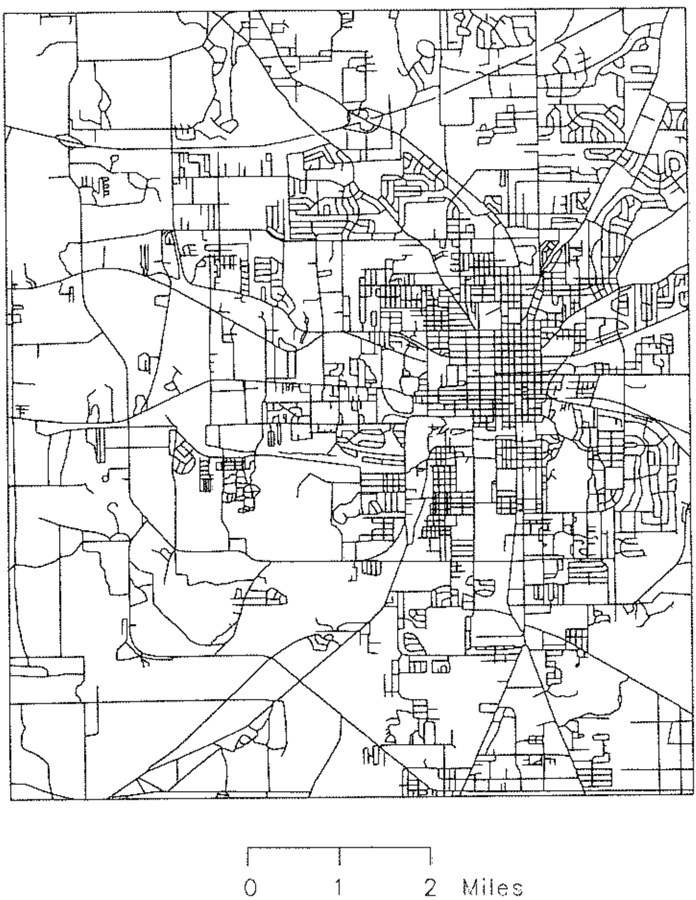


Figure 1.--Plot of digital road network for portion of Tallahassee 1:100,000-scale map

Driving Forces

Technological Factors. A number of design characteristics of the 1:100,000-scale maps were chosen for compatibility with automated digitizing techniques. These characteristics include the feature separation of materials during map compilation, a decrease or simplification of symbolization of linear features (for example, roads represented by solid roads instead of cased roads), and a reduction in the number of features represented by discontinuous lines (for example, intermittent streams are shown by solid instead of dashed lines). Recognition should be given to the designers for their foresight to enhance compatibility with a digitizing system that did not exist at the time.

USGS is attempting to capitalize on these design characteristics and is gathering the bulk of the digital data using automated digitizing techniques. A Scitex Response 250 system is being used to raster scan the feature separates for the roads and stream network. Interactive editing of the raster data is performed on the Scitex and is followed by raster to vector conversion. The vector data are further edited and tagged on an Intergraph edit station, with final structuring to a DLG format occurring on an Amdahl main-frame computer. It is believed that this production system will enable us to produce, at a high rate, large amounts of digital cartographic data. Further details are provided by George M. Callahan in his presentation on "Digital Cartographic Data Production Using Raster Scanning Data" at this symposium (Callahan, 1983).

Programmatic Factors. The fact that USGS has the technical capability and production capacity to create a 1:100,000-scale digital cartographic data base is not sufficient justification to embark on this program. A major validated user requirement making the activity a joint venture was a further inducement to initiate this program. The U.S. Bureau of the Census (Census) has a requirement for digital cartographic data describing roads, railroads, streams, and other geographic features to enable it to automate its geographic support system for the 1990 census. As described by R.W. Marx (1983), Census is developing its Topologically Integrated Geographic Encoding and Reference (TIGER) system to automate many of the geographically based aspects of data collection, processing, and publication activities associated with a decennial census. According to Marx "TIGER will store and identify the hierarchical relationships of all geographic areas represented in each census or survey: the 50 states, 3,137 counties, more than 18,000 minor civil divisions or equivalent areas, more than 20,000 incorporated places, and all other areas . . . TIGER must accurately portray the boundaries of all political and statistical areas recorded in the file, plus the streets, roads, and other map features needed for field operations." The 1:100,000-scale digital cartographic data base will form the cartographic foundation for the TIGER files. USGS and Census have agreed (through a memorandum of agreement) to work jointly in the production of this data base.

Anticipated Uses

The 1:100,000-scale digital cartographic data base will provide the geographic framework for geographic information systems through the rest of the century. The addition of street names and census-keyed geocodes by Census will allow spatial analysis of vast amounts of socioeconomic data. USGS is adding information on land ownership and mineral resources to the base categories to form a data base for resource management decisions on Federal mineral lands (see Kleckner and Anderson, 1983). The 1:100,000-scale digital cartographic data base combines the heretofore elusive qualities of sufficient detail, adequate content, and nationwide coverage that will allow, and perhaps foster, the widespread use of spatial analysis methods. Digital dashboard road maps and automated yellow pages based on the information in this data base are two futuristic concepts that could become reality.

A synergistic effect will be realized through the combination of the cartographic information contained in the 1:100,000-scale data base with digital remote sensing data. The new generation of higher resolution, multispectral remote sensing satellites (Landsat 4 Thematic Mapper, SPOT, and others) are providing detailed data about the Earth's surface. The cartographic data could provide accurate geometric control for the remote sensing data. Networks of linear features (for example, major highways) extant in both data sets could allow for better fit of remote sensing data to a cartographic base than is now possible through the use of isolated control tie points. The digital merging of these two types of data could allow new methods of imagery analysis to be developed. Conversely, if the resolution of the remote sensing data is adequate, direct digital update of features in the cartographic data base may be possible.

SUMMARY

The USGS is beginning a program to create a digital cartographic data base containing transportation and hydrographic features from its 1:100,000-scale maps by the end of the decade. Census is cooperating in the development of this data base and will enhance it through the addition of street names and census geocodes. The existence of this data base will enable the widespread use of geographic information systems for a host of resource management, area analysis, and planning activities. Combined with remotely sensed digital data, our ability to study and monitor the Earth's surface will be improved. This data base will not only meet the immediate (1990) needs of the Federal community, but it may well serve as the catalyst for the widespread use of geographic information systems technology in the United States.

REFERENCES

Allder, W. R. and Elassal, A. A., 1983, Digital line graphs from 1:24,000-scale maps: U.S. Geological Survey Circular 895-C (in press).

Allder, W. R., Beck, F. J., McEwen, R. B., Szeide, A. J., 1983, Digital line graph attribute coding standards: U.S. Geological Survey Circular 895-G (in press).

Callahan, G. M., 1983, Digital cartographic data production using raster scanning data: Sixth International Symposium on Automated Cartography, Ottawa, Canada, Oct. 16-21, 1983.

Kleckner, R. L., and Anderson, K. E., 1983, An integrated data base for energy and minerals information on public lands: Sixth International Symposium on Automated Cartography, Ottawa, Canada, Oct. 16-21, 1983.

Marx, R. W., 1983, Automating census geography for 1990: American Demographics, June 1983, Vol. 5, No. 6, pp. 30-33.

COMPUTER-ASSISTED MAPPING FOR THE CENSUS OF CANADA

D. Ross Bradley
Statistics Canada
Geography Division
Jean-Talon Building
3rd Floor, Section B-7
Ottawa, Ontario
K1A 0T6

ABSTRACT

The Census of Canada which takes place every five years is a huge undertaking. To carry out the Census in an orderly fashion requires a map as an aid to each one of the 35,000 to 40,000 Census Representatives (enumerators), to ensure an exact and complete coverage of questionnaire drop-off. The Census must publish the results of its findings, and since these are made available at a variety of geographic areas, reference maps are a necessity so that their boundaries can be identified. Results are also portrayed in thematic maps, which are provided as an alternative format to the tabular display of data. The computerization of census collection maps, reference maps and thematic maps is well underway at Statistics Canada.

INTRODUCTION

The Census of Canada, often referred to as the largest peacetime operation, carried out by the Government of Canada, takes place once every five years. Its objective is to count all persons living in Canada at their usual place of residence on Census Day. To carry out a task like this requires some 35,000 to 40,000 Census Representatives (enumerators) to drop off questionnaires to about 9,300,000 households during a very short but specific time period. To ensure an orderly coverage, maps must be provided to Census Representatives so that precise geographic areas are covered once and only once throughout the entire land mass of Canada. Thus the need for <u>collection maps</u>. The Census must also publish the results of its findings. Population data are made available at many geographic areas. <u>Reference maps</u> are made available for all these areas. <u>Thematic maps</u> are provided as an alternative format of data display. For the 1981 Census they will be found in a Metropolitan Atlas Series - for 12 large Metropolitan Areas, and will complement the Census Tract Profile bulletins.

This paper describes the computer-assisted mapping for these three types of maps.

COLLECTION MAPS

The need for a huge volume of large scale maps by Census Representatives is evident. Since Census Representatives generally are not trained in map reading, any map to be used must be very clear, precise and easy to read. Producing maps for the census collection activity is costly. For the 1981 Census, this project consumed some 80 clerical and drafting person-years and about $1,500,000. The labour intensive project was studied carefully just prior to the 1981 Census with the view to saving human and financial resources through the use of computer technology, and integration of activities. Thus began the creation of maps for collection purposes by computer-assisted means.

If maps are going to be produced by computer-assisted means there must of course be a machine-readable file in place. There is a file for this purpose, called the Area Master File (AMF) which consists of digitized street network and supporting information such as street names, address ranges and other geographic features. There is an AMF for virtually the entire urbanized core of all centres of 50,000 population or more, which number 37; i.e. from places like North Bay with a population of just over 50,000 to places like Toronto and Montréal which have well over 2,500,000 population in their urbanized core. AMFs contain a logical machine-readable representation of all city streets and other selected features like railroad tracks, rivers, municipal boundaries, etc. The AMFs geographically reference every street, address range, block-face and centroid coordinate in the area covered. A computer-plotted map is an important product of the AMF. Using the utility program MAPMAKR, maps can be produced showing street patterns in a single-line fashion (Figure 1).

Since the collection of data and complete coverage are two of the most fundamental activities of the census, the single-line map as produced from the AMF is considered inadequate for these vital purposes. Therefore, a program was created to produce a double-line street pattern from the single-line network (Figure 2). A prototype system was initiated and some 200 maps were produced for use by Census Representatives during the 1981 Census. Subsequent evaluation of these maps determined that they were more than operationally adequate. Costs of maps produced by computer-assisted means for the prototype were quite similar to costs of maps produced by traditional methods but they were much less labour intensive and the potential for cost improvement was high. During the past year a production system has been in development at Statistics Canada to produce about 8,000 enumeration area maps for the next Census. Costs presently are estimated to be about one-third of what they were for the prototype in 1981. Other benefits are in terms of improved consistency in collection processing and retrieval of map bases.

While these maps have direct application to the Census, other surveys in Statistics Canada, the Labour Force Survey in particular, will find these useful, as will the Canada Post Corporation, the Chief Electoral Officer of Canada and his provincial counterparts.

It is an interesting aside, to note that mapping was not the main reason, initially, for creating Area Master Files. The geocoding system, of which AMFs are the base, was developed in the early '70s to provide complete flexibility for the retrieval, by user-specified areas, of census data assigned to block-face centroids (Figure 1). Collection maps for the census are a by-product of that system, and may very well become the principal use of the AMF.

REFERENCE MAPS

A reference map for census purposes is any map that identifies and locates the boundaries of a geographic area for which census data are made available. Published reference maps are available for census divisions (counties, regional municipalities, regional districts, etc.), census subdivisions (cities, towns, villages, townships, Indian Reserves, etc.), census metropolitan areas, census agglomerations, census tracts, and federal electoral districts. Reference maps such as those shown in Figures 3, 4, 5 have been produced in the traditional manner for many censuses.

As part of the 1981 Census publication program Statistics Canada produced for the first time a series of reference maps with the aid of the computer. Reference Maps for Census Divisions and Census Subdivisions (Catalogue No. 99-907) of the 1981 Census comprise some 41 maps and is the product of computer-assisted cartography. Production of these maps combined automated and manual mapping techniques by integrating three digital files.

(a) A file of census subdivision (municipality) boundaries obtained by digitizing the boundaries from Energy, Mines and Resources Canada base maps at the scale of 1:50,000 and 1:250,000.

(b) The file of names and codes from the Standard Geographic Classification (SGC) of Statistics Canada.

    Note: Both files (a) and (b) were components of the 1981 Census Geographic Master File (CGMF).

    The CGMF includes all the necessary information to enable the extraction of census data for any standard geographic area in Canada. By virtue of its geographic editing capacity this file verifies that each Enumeration Area is coded within its proper geostatistical area.

(c) A file of digitized water-ways and shorelines created from Energy, Mines and Resources Canada base maps at a Scale of 1:2,000,000.

The Universal Transverse Mercator (UTM) boundary file (a) above) was converted to Lambert conformal coordinates and then superimposed on the hydrographic file and overlapping or unnecessary lines were removed interactively on a Cathode Ray Tube (CRT). If a municipal boundary, for instance, protruded into a water-way, it was erased so that only the shoreline showed. Names from the census subdivision file were automatically placed at the centroid of each municipality, linking the geographic code from both files. If names were too crowded they were repositioned interactively for satisfactory spacing. Maps for the Yukon, Northwest Territories, Prince Edward Island, Nova Scotia and New Brunswick were produced using this method (Figure 6). The remaining maps in this series were produced by using more manual intervention and some conventional cartographic operations. The municipal boundaries and water-ways and shoreline files were used with the computer. Unnecessary lines were eliminated by hand rather than via the on-line procedure, while the names and codes were positioned manually, using a computer-generated listing, and a computer prepared sketch, with names located at the centroid of each municipality (Figure 7).

These maps were produced in 3 colours; red for the boundaries of the 266 census divisions and their standard geographic codes; blue for the water-ways and shorelines; and black for the boundaries and the names of the 5,710 subdivisions (and the codes of more than 2,600 when the census subdivision was also a census consolidated subdivision) used for the dissemination of data from the 1981 Census.

THEMATIC MAPS

While most census data are displayed in tabular form and published as bulletins, the technique of displaying statistical data using thematic maps is most effective. As a result of an Ottawa-Hull Census Metropolitan Area prototype Atlas produced following the 1976 Census, it was decided to produce a Metropolitan Atlas Series of the 1981 Census which will depict, in thematic maps and graphic format, the distribution of various characteristics of the population for the Census Metropolitan Areas of Toronto, Montréal, Vancouver, Edmonton, Calgary, Regina, Winnipeg, Hamilton, Ottawa-Hull, Québec, Halifax and St. John's.

Thematic map atlases are widely used in urban planning, social work, and education - primarily to gain general impressions of the data and their geographic distribution that can then be followed up with more detailed study. The Metropolitan Atlases of the 1981 Census will contain a large number of maps and graphs which illustrate a variety of census themes or variables (Figures 8 and 9) for a census metropolitan area. Mapping is at the Census Tract (a small geostatistical area averaging 4,000 persons established in urban communities of 50,000 population and over) level and is designed as a complement to the Census Tract Profile bulletin. It reveals spatial patterns in the data within a metropolitan area not immediately apparent when presented in tabular form.

The Metropolitan Atlas Series is a high technology state of the art product. Significant advances in data analysis procedures, computer mapping software and plotting hardware have been brought together to produce these atlases. Software support for computer mapping is provided by GIMMS (Geographic Information Manipulation and Mapping System) and data analysis of census tract data uses a series of computer programs to assist in the selection of appropriate class intervals and dot rates for the maps and graphics. Each page is produced entirely on a GERBER 4442 high-speed drum plotter which uses a special photohead assembly in which a focussed beam of light passes over a sheet of photographic film mounted on the drum. Pages are printed at actual size so that the film can be sent directly for printing of the bulletins with no intermediate steps or loss of quality. Maps are presented in monochromatic form to facilitate reproduction and because of financial constraints on printing.

The computer-assisted mapping that has been carried out at Statistics Canada has been a cooperative venture of the Geocartographics Sub-division of the Data Processing Division and the Geography Division. The hardware, software and mapping services of the projects described have been provided by the Geocartographics Sub-division. Geocartographics has the expertise to service the geocartographic needs of the census and other areas at Statistics Canada.

In concluding, computer-assisted mapping has made significant advances for the Census at Statistics Canada during the past few years. The consistency that automation brings along with the economies that can be realized far outweigh any disadvantage that a less aesthetically pleasing product brings. There is little doubt that we have barely scratched the surface in this exciting area. Cartographers should take up the challenge and be more aggressive in the field of automated cartography. In so doing the cartographer with sophisticated specifications and the computer scientist with advanced technology can form a team to be envied.

REFERENCES

Bradley, D.R. 1981, Computer-Assisted Collection Maps for the 1981 Census of Canada, pp. 1-5

Geocartographics Sub-division, 1982, Hardware/Software Descriptions, pp. 1-11

Puderer, H.A. 1982, Census Subdivision Reference Maps: Task Documentation

Ross, G. 1983, Metropolitan Atlas Guidebook, pp. 1-25

Statistics Canada, 1972, GRDSR: Facts by Small Areas, pp. 1-21

Wellar, B.S. 1982, Towards a Strategic/Master Plan for Coordinating Geographic, Cartographic and Graphic Census Activities of the Geography Staff, Statistics Canada, pp. 1-46

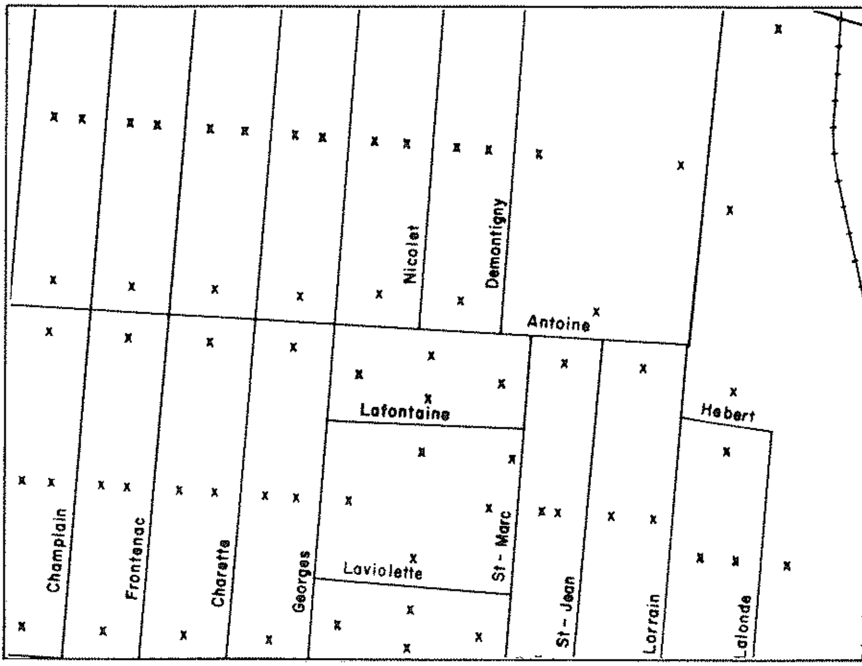


Figure 1. AREA MASTER FILE PRINTOUT (SINGLE LINE)

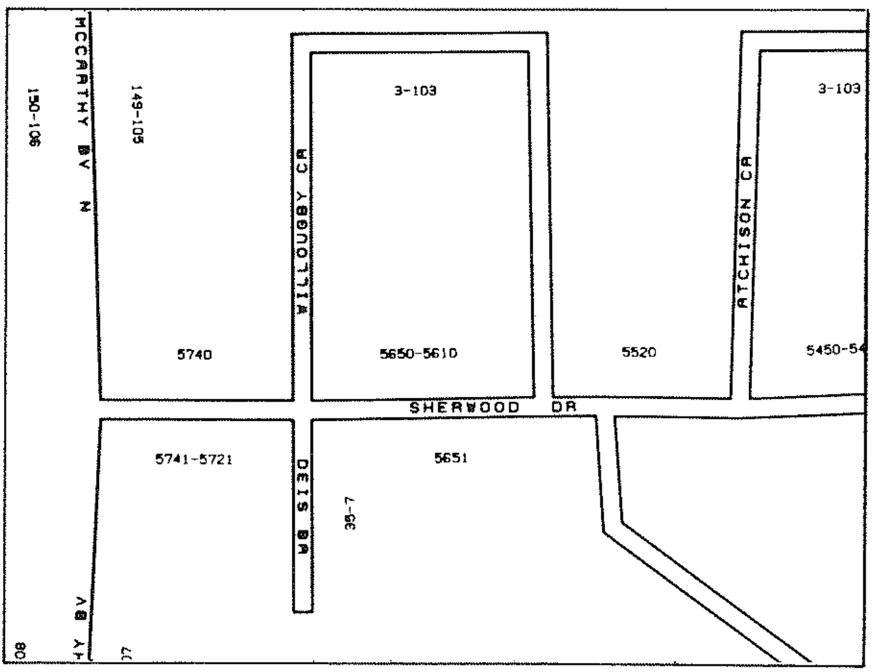


Figure 2. COLLECTION MAP DRAWN BY COMPUTER

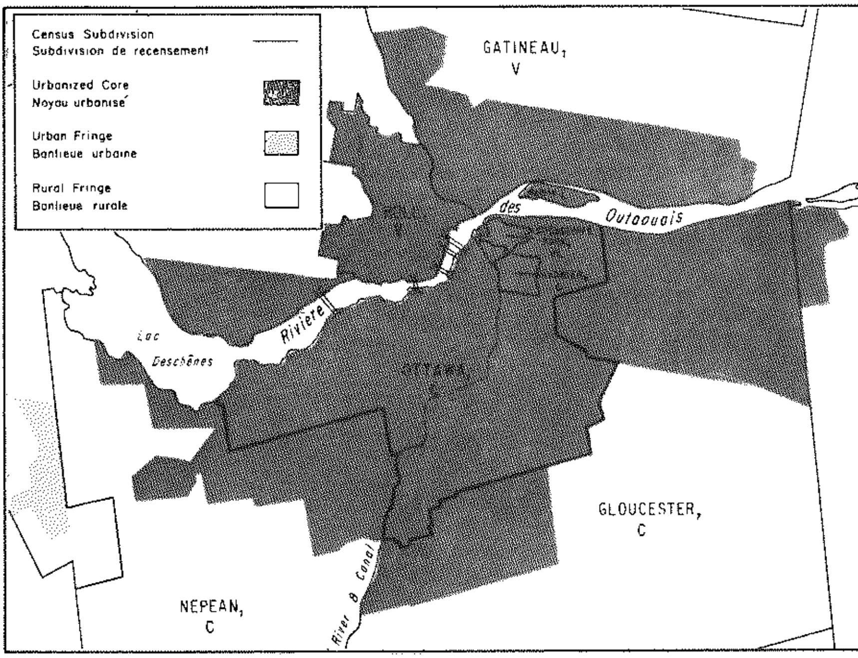


Figure 3. CENSUS METROPOLITAN AREA

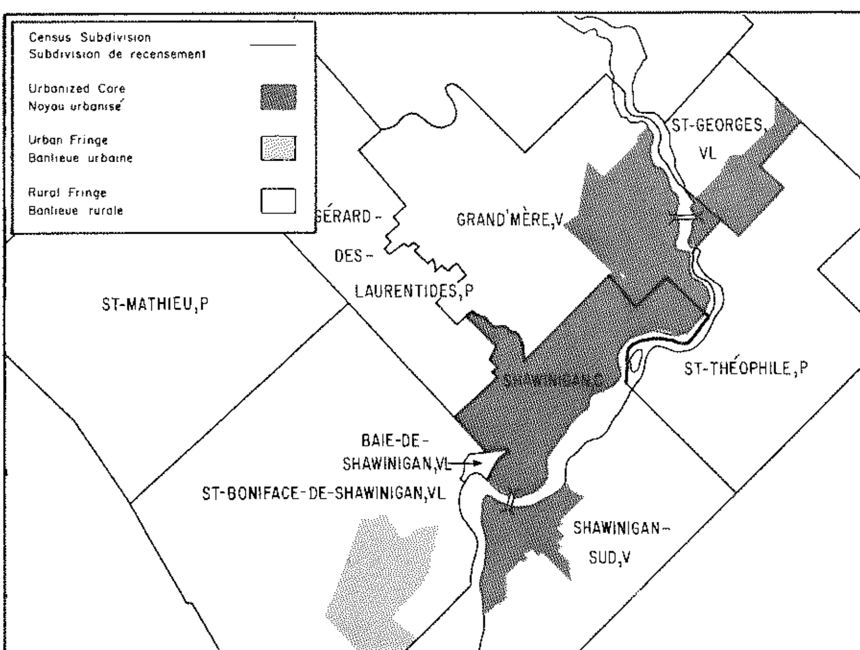


Figure 4. CENSUS AGGLOMERATION

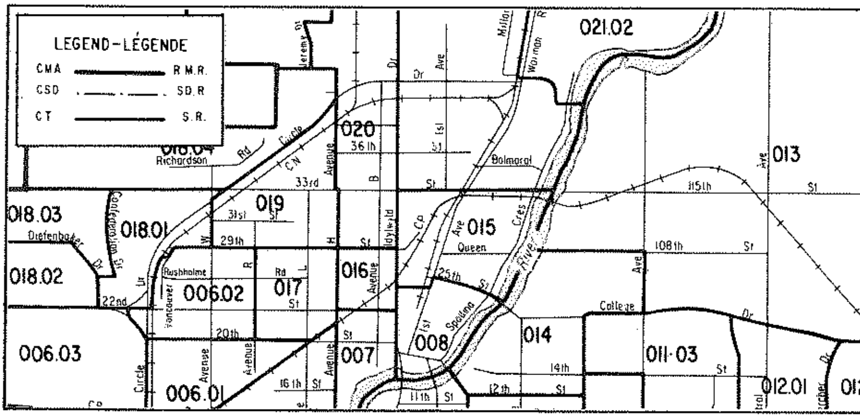


Figure 5. CENSUS TRACT INDEX MAP

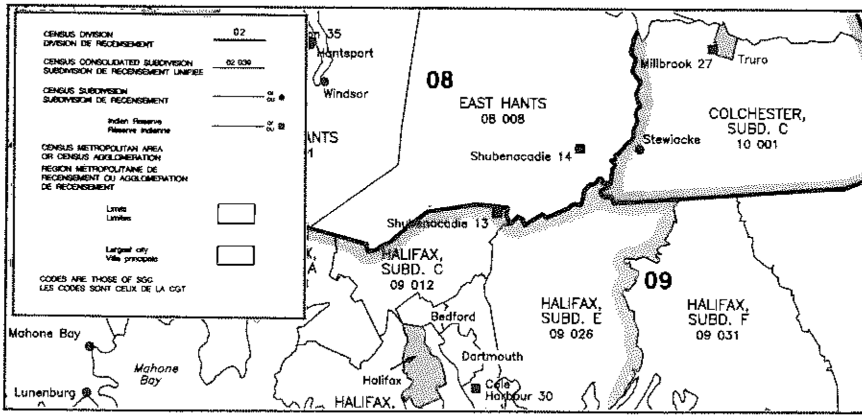


Figure 6. CENSUS SUBDIVISION

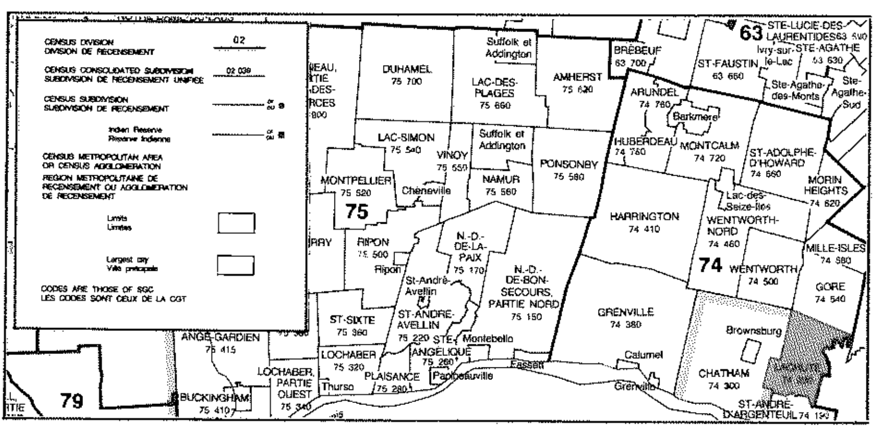


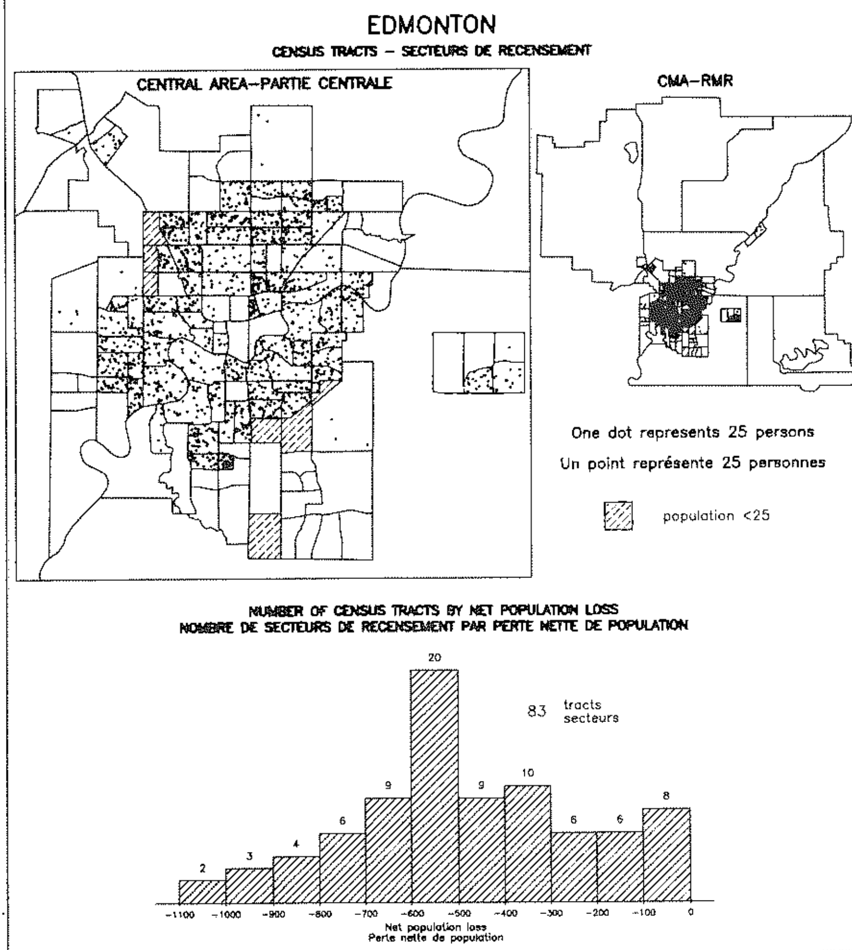
Figure 7. CENSUS SUBDIVISION

NET POPULATION LOSS, 1976 – 1981

Census tracts which experienced a net loss of population between 1976 and 1981 are highlighted. Net losses are shown in absolute numbers of persons.

PERTE NETTE DE POPULATION, 1976 – 1981

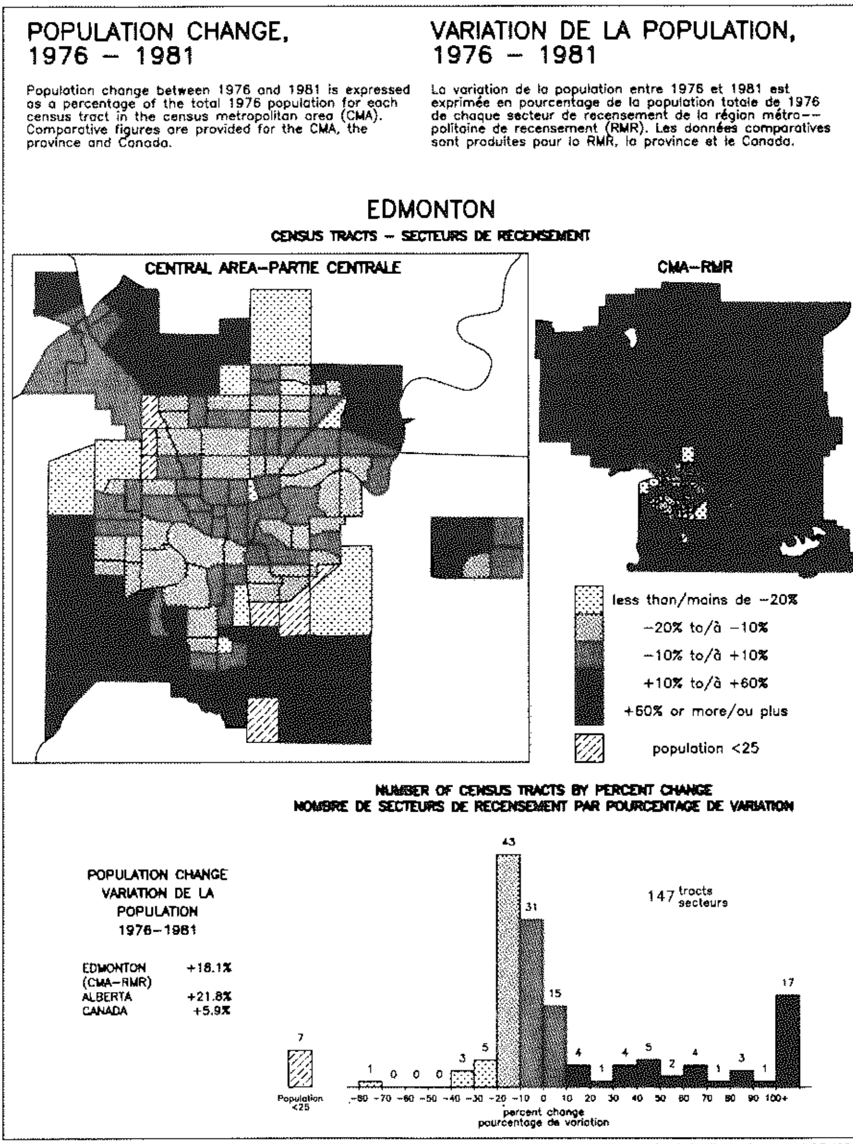
Les secteurs de recensement subissant une perte nette de population entre 1976 et 1981 sont mis en évidence. Les pertes nettes sont données en nombres réels de personnes.



SOURCE: 1981 CENSUS OF CANADA
PRODUCED BY STATISTICS CANADA, 1982.

SOURCE: RECENSEMENT DU CANADA DE 1981
ÉTABLIE PAR STATISTIQUE CANADA, 1982.

Figure 8.



SOURCE: 1981 CENSUS OF CANADA
PRODUCED BY STATISTICS CANADA, 1982.

SOURCE: RECENSEMENT DU CANADA DE 1981
ÉTABLIE PAR STATISTIQUE CANADA, 1982.

Figure 9.

NOTE

Figures 3, 4, 5, 6, 7 are available in colour and can be obtained from the author.

THE AVAILABILITY AND USE OF DIGITAL TOPOGRAPHIC DATA

M. Rodrigue
Research Officer
United States Military Academy
West Point, New York 10996

L. Thompson
Research Officer
United States Military Academy
West Point, New York 10996

ABSTRACT

The difficulty experienced in identifying, positioning and depicting terrain features from spectral digital data has precipitated a need for the photogrammetrically prepared digital data base. Data bases of this type are currently produced by the Defense Mapping Agency (DMA) and the United States Geological Survey (USGS). Because of the highly volatile state of this developmental technology, considerable confusion exists in the minds of the consumer as to what data is at his/her disposal. This paper discusses the types of products available while explaining the content, format and accuracy of each. Issues concerned with reading the tapes are addressed, and several examples are given showing how the data may be employed in topographic and civil engineering problems.

INTRODUCTION

The launching of the Landsat satellites in the early 1970's, combined with the on-going electronic revolution, precipitated tremendous interest in digital topographic data. The literature is replete with tests, analyses, comparisons, and examples of attempts to apply Landsat digital data to various real-world problems. Although Landsat data has been helpful, it generally has not been the panacea for which may topographic and civil engineers hoped.

The 79m pixel size, severe overlap, and the inability to readily derive relief data are the most important shortcomings. In heterogeneous terrain it was impossible to separate small lakes, swamps, rocky outcroppings, and other features which are of interest to the terrain analyst. The civil engineer was similarly unable to directly apply the vast quantities of data at his disposal.

Spectral digital data gathered from airborne scanners, despite its increased resolution, has not rendered itself totally useful. In studies done at the Computer Graphics Laboratory (CGL), Department of Geography and Computer Science, United States Military Academy, it was revealed that, even with a 12.5 foot pixel, a decision algorithm based solely upon spectral response would not be able to routinely identify terrain features. Sufficiently smart software (artificial intelligence) must be compiled to permit the computer to respond not just to the spectral characteristics of the data but to the tonal, textual and spatial characteristics as well.

Unfortunately, it is extremely difficult to start from spectral data and identify, position, and draw the myraid features which are of interest to the terrain analyst and the civil engineer. At this time it has not been established whether it is even possible. Surely more than just the visible and infrared spectral data will have to be available. And even if the visible and infrared bands are combined with radar returns, mid- and thermal-IR responses, and other yet-to-be-determined sources, the transition will be extremely difficult.

It has been demonstrated in the CGL that software (based on texture and shape) can be written to permit the computer to separate trees from other vegetation and an earth road from bare soil. However, despite limited success in these two areas the total problem of adequate scene analysis and feature segregation is recognized as being extremely complicated. Training samples taken from many features which appear to the eye to be homogeneous, and therefore would lend themselves to a probablistic classification algorithm, have been shown to be often not from the same statistical population (Thompson 1982). Seemingly uniform roads, buildings, etc., reflect differently as a result of shadows, sun angle, texture, roof structure, and several other factors. To write software to account for every possibility becomes overwhelming in concept and unbelievably laborious and lengthy in practice.

The previous discussion explains why the photogrammetrically prepared digital data base is becoming increasingly popular. As a minimum it will serve as an interim product for the next 10-15 years.

The following is an overview of some digital products available today. The information contained in these sections was gathered from visits to the agencies' installations. This includes interviews with key personnel and summaries from references listed at the end of this paper. For specific details concerning this information, the authors suggest you contact the agency involved or the references listed.

PHOTOGRAMMETRICALLY PREPARED DIGITAL DATA BASES

Both the Defense Mapping Agency Agency (DMA) and the United States Geological Survey (USGS) are involved in producing digital elevation and feature data.

United States Geologic Survey

The United States Department of the Interior Geological Survey (USGS) National Cartographic Information Center (NCIC) distributes to the public digital cartographic/geographic data files produced by the USGS as part of the National Mapping Program. Digital data files are available in two general categories: elevation data and feature data.

Elevation data. The elevation data distributed by the USGS takes the form of the Digital Elevation Model (DEM) (USER 1978). The DEM consists of a sampled array of elevations for a number of ground positions at regularly spaced intervals. The type of products one might receive from the USGS is dependent upon the accuracy and principal source of the data. A DEM file may be created from existing contour plates, profiles or terrain models scanned in stereo photogrammetric equipment, or from computer-driven orthophoto equipment. The DEM has a raster structure and will accommodate data acquired from many sources.

A DEM file is available at two map scales. The 7.5 minute DEM (USGS produced) uses high-altitude photographs at the nominal scale of 1:78,000 as its principal source of data. The data are resampled to produce one DEM, with a sampling interval of 30 meters for each 7.5 minute, 1:24,000 scale quadrangle.

DEMs of 1:250,000 scale quadrangles were scanned by the Defense Mapping Agency from the contour plates and resampled at a 3-arc-second latitude/longitude interval. The area of each 1:250,000 scale quadrangle has been split from its 1 x 2 degree block to cover a 1 x 1 degree block for each DEM.

A DEM file is organized into a series of three logical records: Type A, Type B, and Type C. The Type A record contains information defining the general characteristics of the DEM. The Type B record contains the elevation data together with associated information. There may be and usually are more than one Type B record organized in patches of coverage. The patches are used where a regular pattern of elevation data points is not feasible over the entire coverage. Each patch will then be regular within and information provided on how they mesh. The Type C record provides the statistics on the accuracy of the data in the file. The files may be obtained in binary or ASCII codes. Physical characteristics of the recorded files will vary depending on user requirements and limitations of the hardware available.

Feature data. The feature data distributed by USGS takes the form of the Digital Line Graph (DLG) or the Land Use/Land Cover Data (LU/LC).

The DLG is the line map information in digital form (DLG 1982). The data files include information on planimetric base categories, such as transportation, hydrography, and boundaries. DLGs are available as 1:2,000,000 scale national Atlas sectional maps and 7.5- and 15-minute topographic quadrangle series maps. The data are recorded in separate overlays or segments if a large amount of data exist for a given section according to major category of cartographic data. These include:

- Political boundaries
- Administrative boundaries
- Roads and trails
- Railroads

- Streams
- Water bodies
- Cultural features (airports, Alaska Pipeline)
- Hypsography (Continental Divide only)

All data contained in the DLG data files are represented on maps as points, lines, and areas. In digital form, these points, lines, and areas are represented as various kinds of lines as follows:

Feature Type	Example	Line Representation
Point	Airfields	Degenerate lines; lines with no length
Line	Roads, railraods, streams	Directly represented as lines
Area	Forests, water bodies	Line describes the boundary of the feature

The data are abstracted and represented as "lines" and are topologically encoded to retain the spatial relationships.

DLGs are available in two data formats: a graphics format (suitable for preparing maps using the GS-CAM software), and a topologically structured format (standard DLG format).

The LU/LC feature data is being compiled on a regular basis by USGS (Hallam 1981). Land Use refers to man's activities which are directly related to the land. Land Cover describes the vegetation, water, natural surface, and artificial constructions of the land surface. These maps are being compiled using a classification system that has a framework of nine general categories that are further subdivided in 37 more specifically defined categories. The general categories include: urban, agricultural, rangeland, forest, water, wetlands, barren, tundra, and perennial snow or ice.

A mapping unit having a minimum size of 10 acres is the smallest area for urban areas, bodies of water, surface mines, quarries, gravel pits, and certain agricultural areas. A minimum of 40 acres is used for all other categories. These units are the smallest areas appropriate for use on 1:250,000 and 1:100,000 scale maps.

Defense Mapping Agency

The Defense Mapping Agency (DMA) likewise produces both digital terrain and feature data. The DMA standard product is currently the Digital Land Mass System (DLMS) (DLMS 1977). The DLMS contains two files, Digital Terrain Elevation Data (DTED) and Digital Feature Analysis Data (DFAD).

Elevation data. There are currently two formats of DTED.

Level I has an absolute accuracy of approximately 30m or 100 ft. The data are formatted such that the spacing is approximately 300 ft. The "standard" product (the preferred one by DMA) is arranged using a geodetic reference system. The spacing is in arc seconds of latitude and longitude. The latitude spacing is always three seconds, but the longitude varies with increasing latitudes to attempt to keep the approximate 300 ft spacing. Level II DTED has the same absolute accuracy as Level I but uses a smaller spacing. A one-second spacing (approximately 30m or 100 ft) is the standard for Level II DTED. In either case a header file identifies the lower left corner of the area coverage and the data interval. Elevation values are stored by increasing steps of longitude within logical records which are themsleves stored by increasing steps of latitude. These data formats are also available in UTM by special request.

Feature data. DFAD formats are numerous and varied. The absolute accuracy of all DFAD is equal to that found on a 1:250,000 map (130m or 425 ft at a 90 percent CEP on the World Geodetic System). Relative accuracy (relationship of one feature to another nearby feature) is much better. The density of data differs for the various levels, but the absolute accuracy remains the same as Level I. Generally, the number of features compiled into the data base varies with the level. The size of the feature is usually the determining factor. For example, a building 30 meters on a side might not be included in Level I but would be in Level II. Level I DFAD is normally delivered with a DTED one degree by one degree cell for a DLMS standard package. These are the most common levels produced.

In addition the following types of DFAD data exist:

- Level IC - used for B-52 navigational aids.
- Level V - for visual. This includes more than just radar reflective features. This is the High Resolution Sensor Simulation data base. It currently exists only in prototype format for limited areas.
- Level X - under development. This format type is to provide cartographic and topographic information about features.
- Level V-HR/TA - This special format of the visual data is designed for higher resolution terrain analysis information. Currently a single prototype data base of the Ft Lewis, Washington area is available.

DFAD format is arranged as a manuscript of features. There are three types of features. As in the USGS DLG data, the feature types include point, line, and area features. Each category is given a feature descriptor used to identify the data relative to the feature. For example, a point feature such as a building might be identified by the digits 100. Since it is a point feature, the data following this code would be its location as relative coordinates based on the reference coordinate given in the header information. A line feature would have a reference number indicating the number of pairs of relative coordinates to follow. An area feature would have the reference number and the pairs (as a check, the last pair would be the same as the first to ensure polygon closure). The descriptors are related by category. If roads were identified by descriptor 200 for example, dual lane roads might be identified by 201 and multilane highways by 202. The new HR/TA data format provides for microdescriptors which are additional digits in the code (HR/TA 1982). For the higher resolution data one would look for 4-digit descriptors in the manuscript. A single area coverage may contain multiple manuscripts.

The newest proposal for DMA digital data is called Digital Data Base 85 (DDB-85) projected for completion in 1985. It is to be the basis for all automated cartography, terrain analysis, etc. It will contain textual material as well as elevation and feature data. DDB-85 is to contain a data element comprehensive enough to satisfy any requirement. Specific users would receive subsets of DDB-85 in DMA's standard format but containing only the information needed to satisfy their particular request. This Universal Transformation Program is under development by DMA, Air Force Aeronautical Systems Division, Army Training and Doctrine Command, and the Naval Training Equipment Center.

APPLICATIONS

Photogrammetrically prepared digital data bases from the DMA are currently being used in the CGL. Because the tape, as purchased, contains more information than one specific user will need, the first step is to list those features which would assist the user in his/her specific task. Without fail elevation data is on everyone's list. Additionally, the military terrain analyst would want to be able to identify and locate water bodies, swamps, all types of vegetation, transportation systems, soil types, built-up areas (industrial, commercial, residential). Because the target use of these data is on a microprocessor, the features are extracted from the tape and stored in separate files on disk. This concept of "separates" is not new to the mapping community and helps to give those somewhat unfamiliar with digital formats a point of reference when later a scene is generated. Later, when the data are transferred to the medium to be used with the microprocessor, the separates will provide the "packaging" needed to overcome on-line storage problems. Once the applicable digital information has been transferred from tape to files on a disk the user may quickly recall features, overlay them onto the elevation, combine features and, in general, commence a meaningful comparative and analytical study. At this point the uses for the data are bounded only by the imagination and software compiling ability of the user.

Military Applications

An obvious application would be optimum route selection. It may be a requirement for foot soldiers to move from A to B with the following requirements:

- minimize the possibility of aerial detection,
- navigable slope must be less than 50%,
- avoid all swamps, built-up areas and rivers wider than

200 feet, and

- maintain line-of-sight radio communication at all times with point C.

Using DMA's Level V HR/TA digital data the solution to problems of this type are readily manageable. Currently under development in the CGL is a Division Level Terrain Analysis (DLTA) station. This microprocessor-based system is designed to provide the engineer terrain analysis teams in the divisional intelligence staff section with a means to provide rapid information to a division commander for tactical operations. The typical functions required include cataloging mountains of textual reference material, processing digital terrain data, creating numerous tactical overlays (cross country movement, lines of communication, zones of entry), and generating reports to be incorporated in division operations orders. Using the separates as mentioned above, the analyst may then evaluate each of the contributions that particular feature groups may make toward the development of the required product.

The determination of fields of fire or line of sight for radio communications may be readily computed from the DTED by applying well known software developed at the Engineer Topographic Laboratories and the CGL.

Selections of optimum landing zones based on such variables as elevation, distance, size, approach obstacles, enemy positions, etc., are easily made with the algorithm addressing the various digital files and placing the information at the disposal of the planner.

Civilian Applications

Civil engineers may also enjoy the speed and flexibility provided by digital data if they utilize the Geological Surveys' DEM's and DLG's.

First-estimate cut and fill could be quickly obtained from DEM's which have a vertical RMSE of 7 feet and a horizontal RMSE less than 10 feet (User 1978).

Flood control or damage estimates could be greatly facilitated through the correct combination of Land Use/Land Cover digital data (shows urban and residential areas), DLG's (bridges, roads, boundaries) and DEM's. For example a river could be permitted to digitally rise 10 feet above its normal level and the computer instructed to color in blue all areas which would then be under water. The water level could be varied and studies performed to arrive at various damage assessments. The impact of dams and leevees could be assessed prior to construction and the optimum location chosen at the computer in a matter of minutes.

Using the USGS node attribute concept optimum routee may be planned for moving extra wide or heavy loads, dangerous cargo, etc. The steepness of any stretch along a road, the proximity of residential or urban areas, distance, type of road, and many more variables are provided in digital form and could be included in a decision algorithm.

The Land Use/Land Cover digital data combined with the boundaries provided in the DLG's would allow a careful analysis of zoning laws or land use changes.

CONCLUSIONS

The digital data being routinely supplied by the USGS and DMA can be of tremendous value to the topographic and civil engineers. The technology of tomorrow promises real-time analysis from satellite-based spectral sources. Until then, photogrammetrically prepared digital data bases will assist in automating engineer functions. While the data available today lacks the detail for some functions (terrain analysis, for example), efforts are improving at a rapid pace. The computer storage problem remains a key issue, but it also is being improved by rapidly developing technology. The data currently available are adequate for todays needs and for development of future systems.

REFERENCES

"Digital Line Graphs from 1:2,000,000-Scale Maps", User Guide/Interim Edition, United States Geological Survey Publication, 1982.

Hallam, C.A. 1981, "A Guide to Land Use and Land Cover Digital Data in the Geographic Information Retrieval and Analysis System Format", USGS.

"Product Specifications for Digital Land Mass System (DLMS) Data Base", Defense Mapping Agency, 1977.

"Prototype Product Specification to Support High Resolution/Terrain Analysis Applications", Defense Mapping Agency, 1982.

Thompson, L.G. 1982, "A Statistical View of Automated Classification", Paper submitted for publication, Photogrammetric Engineering and Remote Sensing.

"User Guide to Digital Elevation Models", A draft United States Geological Survey publication, 1978.

TWO WAY DATA TRANSFER BETWEEN ARIES IMAGE ANALYSIS SYSTEM AND ARC-INFO GEOGRAPHIC INFORMATION SYSTEM

Robert V. Maher, David Colville and Douglas Rigby
Department of Computer Programming
Nova Scotia Land Survey Institute
Lawrencetown, Annapolis County, Nova Scotia BOS 1M0

ABSTRACT

Nova Scotia Land Survey Institute has Aries II image analysis software from DIPIX Systems Ltd. residing on a LSI 11/23 processor and also ARC-INFO geographic information system from ESRI residing on a PRIME 550 processor. These computing resources support diploma programs in Remote Sensing and Scientific Computer Programming. To better utilize the computer resource, software has been developed to permit transfer of data on magnetic tape between the two systems. The data transfer is effected in raster format, although the ARC-INFO system allows conversion between polygon and raster file formats. This two way transfer has several advantages: it permits use of input and output devices on both processors; it allows analysis of map data with satellite data on either processor using the software tools of that environment. Educationally, it has developed an understanding of the file structures in the Aries II and ARC-INFO systems within their respective hardware environment.

INTRODUCTION

Vertical integration of cartographic data and remotely sensed information about the earth's surface for purposes of resource management has been well recognized by government and the computer industry. At the federal level the concept underlies the desire for standard data transfer formats between Canada Centre for Remote Sensing (CCRS) and Canada Land Data Systems (CLDS), plus other departments, namely Statistics Canada and Agriculture Canada. Provincially, within Alberta the use of the POLYGRID* system has demonstrated the requirement for the same capability (Niemann and Langford 1983, Maher, Boyd and Langford 1980). Within the Canadian computer industry DIPIX has recently announced an agreement with Intergraph, and at an earlier date Environmental Systems Research Instituted integrated their software with the ELAS system from NASA.

While the underlying concept is well understood there tends to be a paucity of information on the technical aspects of the integration of these data types and secondly, the ramifications in terms of specific applications and their associated methodologies.

During the last year, the Nova Scotia Land Survey Institute

*POLYGRID. Acronym for a suite of GIS programs developed by Geospatial Research Corporation, Edmonton. Software used by variety of clients, primarily in Alberta.

(NSLSI) has acquired both a stand alone Aries II image analysis system from DIPIX Systems Ltd. and the ARC-INFO geographic information system from ESRI**. Given diploma programs in Scientific Computer Programming (SCP) and Remote Sensing it was incumbent on the senior author to establish effective data transfer between the two software systems running on different vendor hardware. As co-operative projects within the SCP program, the junior authors developed the programs and documented the in-house procedures for the two way transfer of data. These procedures have been tested by course assignments in Geographic Information Systems at the Institute and student projects within the Remote Sensing program. The specifications of this two way data transfer represents the larger first part of the paper. The second part gives a preliminary description of a research project developed in relation to the fruitland regions of Canada studies by CLDS. In this study, the objective is to develop a methodology based on vertical integration and to explore the inherent limitations of this form of analysis.

	Aries II image analysis system (DEC environment)	ARC-INFO geographic information system (PRIME environment)	
	HARDWARE		
CPU	LSI 11/23	PRIME 550	
Memory	112K	1 megabyte	
Disk capacity	200 mb	160 mb	
Printer	Paper Tiger	Printronix	
Tape	Kennedy 9100	Kennedy 9100	
Video display subsystem	Hitachi colour monitor, Video memory, Houston Instrument digitizing pad, LSI 11/02	Digitizing subsystem	Calcomp/Talos 8000 digitizer Nicolet/Zeta drum plotter
Console	VGT100	Programmers Terminals	36 Volker Craig 414 video display terminals
	SOFTWARE		
Operating System	RSX-11M	PRIMOS	
Languages	Fortran IVP Macro 11	Fortran IV, Fortran 77, PMA, PASCAL, COBOL, BASIC, PL1	
Applications Software	Aries II image analysis system	ESRI: GRID, PIOS, ARC-INFO. DBMS, INFO, IGL, SPSS, POLYGRID, HOLOSONICS, CPL	

Table 1. Computer resources at N.S. Land Survey Institute

Table 1 summarizes the main characteristics of the two

**Reference to specific vendors of hardware and software is for purposes of factual clarity. It should not be construed as unqualified endorsement of a product nor as an in depth critical evaluation of the marketplace.

computer systems at NSLSI. There are several noteworthy points:
(1) the image analysis system and the GIS software operate under different operating systems on different vendors hardware;
(2) the LSI 11/23 does not include a hard copy output device (if we reject the Paper Tiger) whereas the PRIME has two PRINTRONIX matrix printers which can be addressed in plot mode and the Nicolet/Zeta plotter;
(3) the color monitor component of the video display sub-system exists only on the LSI 11/23, likewise the full size digitizing table and the 36" drum plotter are located on the PRIME 550.

Regardless of the different analytic capabilities of the two software systems, the establishment of data transfer procedures would be justified simply in terms of rationalizing input and output devices.

The capability of detailed digitizing of map coverages on the CALCOMP/TALOS 8000 is an asset to the Aries II system at the input stage. At the display stage, it is advantageous to be able to produce printer maps and plots of image files generated by the Aries II software. Conversely, it is an asset to produce colour displays of image files generated using the ESRI/GIS software.

DIPIX TO PRIME TRANSFER
(Aries II file format to ESRI/GIS file format)

DIPIX distinguish between image and non-image files. An image file is in raster format, stored line by line representing a contiguous geographic area.

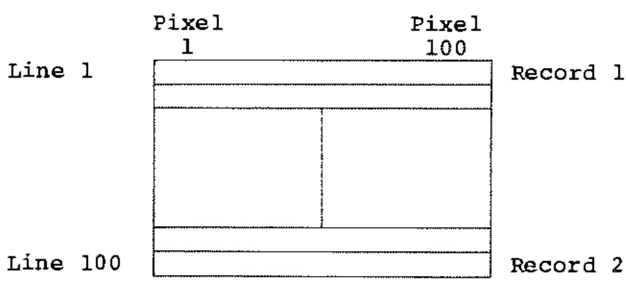


Figure 1. Image file structure on DIPIX system

Figure 1 depicts an image file for an area of 100 lines x 100 pixels. The size of the pixel, as it appears on earth's surface, depends on the data source e.g. LANDSAT, airborne MSS, MEIS, etc. A non-image file may contain statistics or a list of line/pixel numbers for a particular training area. Our focus is the transfer of image files. At NSLSI there are other programs to transfer non-image files to the PRIME.

An image file can be either theme or feature. A feature image file is normally raw satellite or airborne MSS data and represents reflectance values within a range of 0-255. The data is stored as a binary value in a single byte. A theme image file is normally the result of a classification on the Aries II system. In general, depending on the class of a particular pixel, a single theme bit is set. Consequently, within a single byte, the decimal equivalent is either 1, 2, 4, 8, 16, 32, 64 or 128, that is, the bit which is set corresponds to 2^n where n is the bit position between 0 and 7. There are exceptions to this statement: if the user uses the parallelopiped classifier and obtains multiple-classified pixels more than one bit may be set or the user employs the same theme file for the results of more than one classification.

Before describing the details of the procedure, it is worthwhile to identify the three main steps (Rigby 1983):
(1) write image file from disk to magnetic tape in the DEC environment
(2) read magnetic tape onto disk in the PRIME environment converting from DEC bytes to PRIME 16-bit words
(3) reformat disk file for input into ESRI/GIS system

Step 1
Under RSX-11M, there exist system utilities for non-image file transfer e.g. FLX. Within the Aries II system the task BR (Backup and Restore) permits DIPIX users to transfer data between Aries II installations. For the purposes of creating a magnetic tape of an image file for transfer to a non-DEC, non-DIPIX environment, the task TO (Tape Output) was used. While this allows transfer of a single file at a time, it had the advantage of a simple file structure on the tape, whereas BR, for example, includes directory blocks and is designed for transfer of any mix of image and non-image files.

Step 2
On the PRIME, a separate program was written to handle tape reading, byte to word conversion and writing of the disk file. Again, the decision was made to reject systems utilities on the PRIME since they were more flexible but cumbersome and the requirement was a well documented program specific to this task.

Step 3
Environmental Systems Research Institute (ESRI) have developed several GIS systems over the last decade. At NSLSI, we have three of their systems: PIOS (Polygon Information Overlay System), GRID and ARC-INFO. Briefly PIOS, as its name indicates, is a polygon based GIS system, whereas GRID is a grid (or raster) based GIS system. The latest system ARC-INFO replaces PIOS; besides the polygon manipulation routines it incorporates arc digitizing and uses relational database structure of INFO for report generation. Between these three systems exist software interfaces for reformatting and file transfer.

The third step in the DIPIX-PRIME transfer was a program to reformat the image file from DIPIX on disk for input into the GRID GIS system. This was achieved by creating a file compatible with FILMOD. (FILMOD is a program for bringing data files which have been manually created by the user into the GRID system).

Within the GRID system, routines exist for mapping on the PRINTRONIX printer (GRDMDL) or on the Nicolet/Zeta plotter (GRIDPL). Of greater significance are the modelling and analysis capabilities within the GRID environment.

PRIME TO DIPIX TRANSFER
(ESRI/GIS file format to Aries II file format)

Within the ESRI/GIS environment it is possible to digitize polygon, line and point coverages. The digitizing procedure creates a series of arcs which can be built into polygon, line or point features; a set of spatial attributes concerning these features are available to the user through INFO. Within the ARC-INFO environment there are groups of commands for the display, overlay, modelling and analysis of each coverage type and their interrelationship. Similarly, there are commands for the conversion of polygon, line and point coverages into a grid file structure compatible with the GRID system. Given this capability it is possible to identify four steps in the transfer procedure (Colville 1983).
(1) convert digitized polygon, line or point coverage into a single variable file (SVF) in the GRID system using the commands POLYGRID, LINEGRID or POINTGRID
(2) read SVF file from disk and write to magnetic tape in a format compatible with DIPIX system
(3) read magnetic tape on DIPIX system using tasks TF (Tape Format) and TI (Tape Input) to create feature image files
(4) modify header records of the image file from feature to theme.

Step 1
Under ARC-INFO, the three commands, POLYGRID, LINEGRID and POINTGRID rasterize the appropriate cover type. There exists considerable flexibility and hence user options with each of these commands. These options address, in particular, questions related to the presence of multiple polygons or other features within a single grid and the priorities for resolution of this conflict.

Step 2
At this stage, we have a disk file with each pixel having an assigned class value. The program on the PRIME permits the user to specify the relationship between a code value in the SVF and the bit setting for the data transfer to the DIPIX system.

On the DIPIX system, the colour monitor can be used to colour code pixels only from a theme image file. A theme image file contains pixels where within a byte only one bit is set on. Consequently within a byte it is possible to represent one of eight classes, plus a ninth class when no bit is set on, indicating an unclassified pixel. The maximum number of colours available on the monitor is thirty-two. This means that the task for displaying themes on the monitor can handle four theme files at once. Thirty-three classes (including unclassified class) is currently a design constraint on the use of the colour monitor for map display.

If the SVF contains less than or equal to eight separate classes, it is a simple process to develop a look-up table between each class code and its corresponding theme bit setting. If the SVF contains nine to sixteen separate classes then the first theme file will match the first eight codes with theme bits 1-8, whereas the second theme file will match the next eight codes with theme bits 1-8 on that file. The same logic can be expanded to the full thirty-three classes. Since any pixel can be assigned only one code on the SVF if a pixel is set on the first theme file, then that pixel location will be set to zero on all subsequent theme files. The converse is true for those pixels set on theme files two, three and four (see diagram).

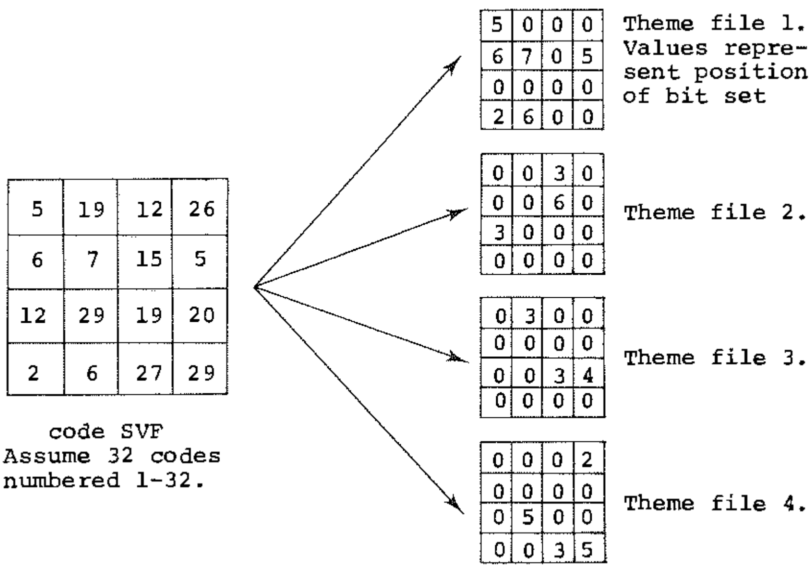


Figure 2. Mapping of class codes in SVF into four theme files

Depending on the number of classes, from one to four band interleaved files will be written to the magnetic tape.

<u>Step 3</u>
The DIPIX tasks TF (Tape Format) and TI (Tape Input) were used to read the files from magnetic tape and write them to the image disk. This task is designed to read feature files (that is, raw reflectance between o-255) however there is no interrogation or modification of the values within the individual byte.

<u>Step 4</u>
Using option 9 of the task DIR, the user changes the header

description of the transferred files from (FF) feature to (TT) theme image files. At this stage, the map can be displayed on the colour monitor using task ID (Image Display).

APPLICATION OF TWO WAY TRANSFER

Display

The ability to display cover type classifications or geological lineaments derived from remotely sensed imagery on the matrix printer or drum plotter has already proven useful for several projects. These facilities are beneficial in so far as the display software can provide map data at the same scale, and registered to the same coordinate system. This allows for easy interpretation between data sources. Furthermore the ARC-INFO system provides various spatial attributes (e.g. area, perimeter and length) which permits comparison of classification statistics.

The transfer of feature image files (raw reflectance values) is less critical at NSLSI since we have other programs which directly read raw data tapes into the PRIME, indeed we use alternative mapping software for producing grey tone printer maps.

The opportunity to transfer map images to the DIPIX system for display purposes is less significant, particularly if the colour monitor is regarded as the primary output device. However there are other high cost output devices which are available for the DIPIX. Given the creation of a DIPIX compatible file other tasks (e.g. OA Applicon plotting, PL Trilog Colour Printer/Plotter or laser film recorder) could produce high quality products.

Analysis

It is within a research context that the data transfer procedures have been developed at NSLSI. At this time, a contract exists with CLDS to produce land use maps of the Annapolis fruitlands at ten year intervals (1961, 1971, 1981), at a scale of 1:50,000, from aerial photo interpretation. A pilot study area (18 km x 16 km) was selected around the Wolfville area for 1981. The polygon coverage was digitized under the ARC-INFO system. It includes 527 land use polygons, some of which are less than 1/8" x 1/8" on the map. Given, a cleanly digitized coverage and having created a code file of land use cover and activity, it was possible to obtain descriptive statistics about the size, number and type of each polygon.

The second stage in the analysis has been to convert from polygon to grid, where the grid is registered to UTM and the cell size is 50 m x 50 m.

On the DIPIX system, it was possible to select a DICS corrected image for the same geographic area and for the same data. The task AP (subarea creation) was used to create an image file which corresponded precisely with the study area. This image was classified, at this time, using an unsupervised classification into eight classes.

From the DIPIX system, the four feature image files (raw reflectance data) and one theme image file were transferred into the ESRI/GIS system on the PRIME.

Within the GRID environment, it is possible to create a multi-variable file which, in concept, has the following structure.

Line #	Pixel	Band 4	Band 5	Band 6	Band 7	Map Class	Image Class
1	1						
.	.						
.	.						
.	.						
1	n						
2	1						
.	.						
.	.						
.	.						
.	.						
.	.						
.	.						
m	n						

Figure 3. Table structure for Annapolis fruitlands in GRID system

With this file structure, it is possible to ask a host of research questions.

(1) Can a confusion matrix be created between the map class and image class?
(2) For those pixels classified as orchard on the map class what are the histograms on all four bands?
(3) How good is the registration between the rasterized map data and the corrected image data?
(4) At what scale do orchards become impossible to distinguish using satellite data?

Under the GRID system, it is possible to develop models (essentially a set of logical relations between variables) and to produce hard copy maps (e.g. the program GRDMDL). Alternatively, the file can be linked into INFO system and summary reports can be generated using its programming language.

At this time, there has been no effort to transfer from the GRID system into the ARC-INFO environment. The transfer software GRIDARC has not been released by ESRI. However, even if this routine was available, it is questionable whether the nature of rasterized LANDSAT data would justify this conversion. Langford (1983) has found that the very nature of LANDSAT raw data does not lend itself to conversion from grid to polygon. Thematic data may be more suitable, especially after spatial filtering however this may negate the original research objective.

CONCLUSION

Our focus has been on the technical aspects of transferring data between an image analysis and a GIS environment. While the DIPIX (DEC) - ESRI (PRIME) linkage is specific to this Institute, increasingly, DIPIX are making software available on the VAX system, likewise, ESRI have converted their software to run in the same environment. In the future the requirement will be the understanding of the different file structures within such environment.

At this time, DIPIX have tasks (DF and DN) which permit digitizing of map data. These routines are used for the identification of ground control points for map to image registration and also to identify training areas. The establishment of data bases for sophisticated manipulation within a GIS system requires continued assessment of remotely sensed imagery in terms of spatial resolution and positional accuracy. These aspects are particularly critical if the information is intended for map updating purposes.

At NSLSI, the DIPIX system permits the display and classification of various types of imagery: LANDSAT, NOAA, RADARSAT, airborne MSS and MEIS. To identify the significant properties of each type of imagery for justified inclusion within a geographic information system represents the future challenge.

Two way data transfer between the image analysis and geographic information system environment is an essential first step.

ACKNOWLEDGMENTS

The authors wish to thank the software specialists at DIPIX Systems Ltd. and ESRI for clarification on the programming aspects of their respective systems. Jeff Freeman and Denise McCullough made available a manuscript map derived from aerial photo interpretation of the Wolfville area from their Annapolis Fruitlands contract with CLDS.

REFERENCES

Colville, D. 1983. Data transfer between ARC-INFO geographic information system and Aries II image analysis system. Final report submitted for Diploma in Technology, Scientific Computer Programming, Nova Scotia Land Survey Institute, Lawrencetown, Nova Scotia.

DIPIX Systems Ltd. 1983. Aries-II Application software package: operators manual. Ottawa, Ontario.

ESRI. 1983. ARC/INFO Users Manual. Version 2.10. Redlands, California.

Langford, G. 1983. Personal communication.

Maher, R.V., M. Boyd and G. Langford. 1980. Integration of geographic information and LANDSAT imagery: a case study of vegetation mapping in the White Goat wilderness area, Alberta. Canadian Journal of Remote Sensing 6(2) pp. 86-92.

Niemann, O., G. Langford and G. More. 1983. Avalanche hazard mapping integrating LANDSAT digital data and digital topographic data. Paper presented at 9th Canadian Remote Sensing Symposium, Montreal, Quebec.

Rigby, D. 1983. Data transfer between Aries II image analysis system and ARC-INFO geographic information system. Final report submitted for Diploma in Technology, Scientific Computer Programming, Nova Scotia Land Survey Institute, Lawrencetown, Nova Scotia.

ON THE TRANSFER OF REMOTE SENSING CLASSIFICATIONS INTO POLYGON GEOCODED DATA BASES IN CANADA

D.G. Goodenough
G.W. Plunkett
Canada Centre for Remote Sensing
2464 Sheffield Rd.
Ottawa, Canada
KlA OY7

J.J. Palimaka
Intera Environmental Consultants Ltd.
Canada Centre for Remote Sensing
Ottawa, Canada

ABSTRACT

In order to facilitate the updating of geocoded data bases with timely resource information, a standard format for polygon data was defined in 1979 through the efforts of the Spatial Data Transfer Committee (SDTC) (Goodenough et al., 1979). This format has been adopted as a standard by four Canadian federal government departments, the LANDSAT Ground Station Operations Working Group (LGSOWG) and several provincial agencies in Canada. One of the objectives of the SDTC was to demonstrate the two-way transfer of data between the Canada Centre for Remote Sensing (CCRS) of Energy, Mines and Resources Canada and Lands Directorate of Environment Canada. This paper describes the efforts to accomplish that objective, notes the successful rasterization at CCRS of polygon data from Lands Directorate and explains the remaining topological problem which prevents the generation of thematic maps from the data after it is entered into the Canada Geographic Information System (CGIS) data base at Lands Directorate. The limitations of the software developed for this application are described and our future plans in this area are outlined. The successful transfer of SDTC Format data from CCRS to Statistics Canada, which is fully described in a companion paper (Korporal, 1983), is noted.

INTRODUCTION

In order for a data base of geographic information to be useful for purposes of planning and resource inventory it must be periodically updated. An economical way to achieve this currency may be through the operational use of remotely sensed data. Geographic information is commonly stored as sets of points defining the boundaries of polygons along with the attributes of these polygons. Data from the image analysis systems used to process remotely sensed images is normally in grid or raster form. Therefore, the first two issues to be addressed before classified satellite and airborne imagery can be used to update a geographic information system (GIS) are:

1) standards must be established for the communication of polygon data in digital form from various information sources into the GIS, and
2) systems must be developed to convert grid data into a polygon format.

The Spatial Data Transfer Committee was set up by volunteers from four agencies within the Government of Canada to consider economical solutions to the problems inherent in the updating of geocoded data bases. The members of the committee were from the Canada Centre for Remote Sensing (CCRS) of the Department of Energy, Mines and Resources, the Geocartographics Group of Statistics Canada, Lands Directorate of Environment Canada and the Lands Resource and Research Institute of Agriculture Canada. The objectives of the committee were:
1) to devise a standard format for the interchange of spatial data,
2) to develop software to read and write standard format tapes, and
3) to conduct a demonstration data transfer.

The first objective was fulfilled with the production of a report defining a standard format for the transfer of geocoded polygon data (Goodenough et al., 1979). The SDTC format has been widely adopted in Canada and has been accepted as a standard by the LANDSAT Ground Station Operations Working Group (LGSOWG) although it has not yet been implemented in an operational system.
This paper describes efforts at CCRS to accomplish the second and third objectives of the committee, namely; the implementation of software to convert remote sensing data from raster to polygon form and to produce SDTC Format data tapes which could be entered into the Canada Geographic Information System (CGIS) data base at Lands Directorate; and the implementation of software to convert data from the CGIS into raster form for use by image analysis systems at CCRS. Related work has been carried out at Lands Directorate to reformat CGIS data to SDTC Format and to read SDTC Format data into the CGIS. That work is not described in detail here.

THE SDTC FORMAT

The SDTC Format was designed to have the following properties:
1) machine independence - the format should be processable on a variety of computer systems;
2) expandability - allowance should be made for revision and extension of the format;
3) generality and flexibility - the format should be adaptable to a variety of geocoded data types;
4) self-definition - the tape format should be self-defining.

To satisfy the requirement of self-definition the format incorporates superstructure concepts and strives to conform to the design standards for Computer Compatible Tape Formats as established by the LANDSAT Ground Station Operations Working Group (LGSOWG, 1979). The superstructure is a combination of precisely defined records and a method of employing them. In general, a tape can be made to conform to the superstructure standard by adding a Volume Directory File to the beginning of the tape and adding file descriptor records to the beginning of each file on the tape.

The Volume Directory File

Figure 1 shows the tape file organization adopted for the SDTC Format within the superstructure scheme. The first file on the tape is the Volume Directory File which describes the logical volume and contains information about the types of files on the logical volume. If a logical volume spans more than one physical volume, each physical volume begins with a Volume Directory File. The first record in the file is the Volume Descriptor Record which describes the logical volume and gives labelling information on the physical volume. Following the Volume Descriptor Record are the File Pointer Records, one for each file on the tape, excluding the Null Volume Directory File.

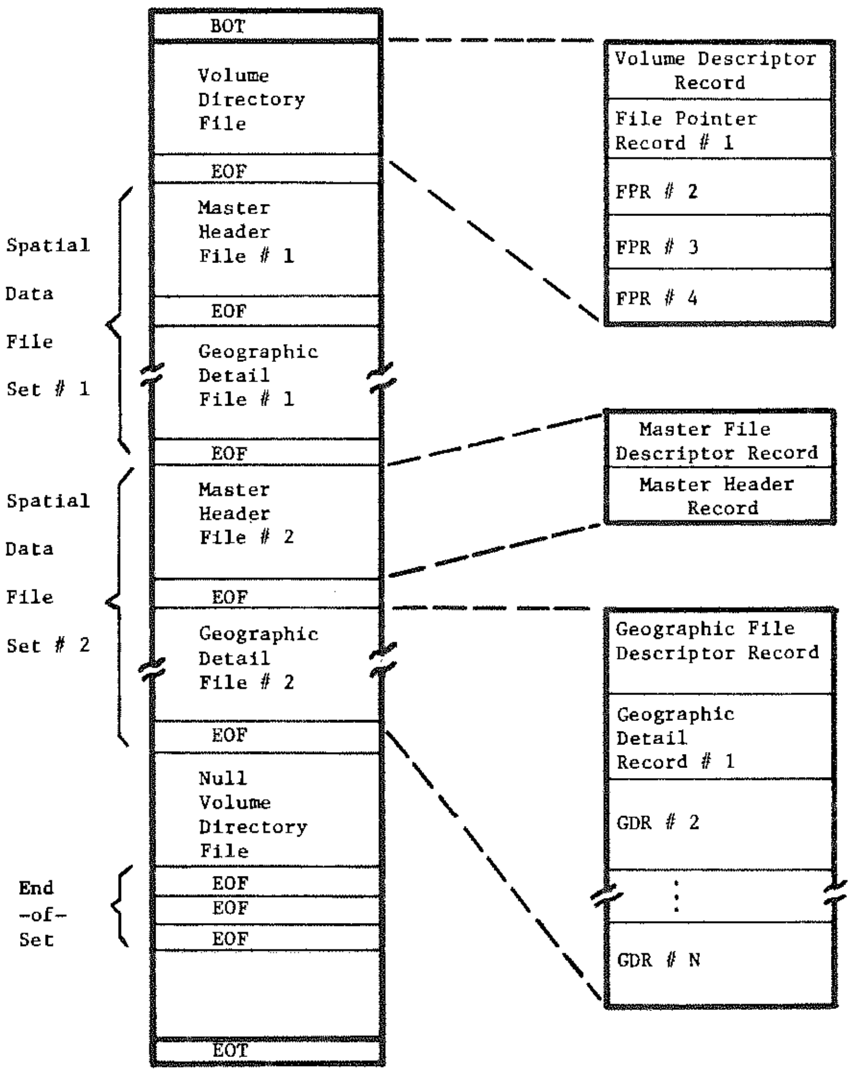


Figure 1. SDTC Format Tape Organization (from Goodenough et al., 1979)

<u>The Spatial Data File Set</u>

The Spatial Data File Set consists of a Master Header File and the Geographic Detail File. The Master Header File contains information on the type of encoding employed to store the data and geographic reference points for the data on the tape. The Geographic Detail File contains the geocoded information. The implementation used for geocoded polygon data is referred to as the Chain Data File which contains the polygon data in the form of chains (defined below).

Normally, an SDTC Format tape will have one Spatial Data File Set and therefore, there will only be two File Pointer Records in the Volume Directory File.

The Null Volume Directory terminates a logical volume. It consists of one record identifying the file as the Null Volume Directory and is normally followed by the end of volume marker.

The records in all of the above files consist of mixed 7-bit ASCII in 8-bit bytes and 32-bit binary integers in 4-byte fields. Full details on the interpretation and data type of each field in each record are given in Goodenough et al. (1979).

The Chain Data File

The Chain Data File in the Spatial Data File Set contains the polygon information in the form of strings of UTM coordinates running from one node to another and making up the common boundary between two polygons. This string of points is called a chain. (A node is defined as a point where three or more chains terminate.) The Chain Data File consists of a file descriptor record and any number of Chain Data Records. Each record is a multiple of 180 bytes in length. Each record contains a chain number and a subchain number. Each chain also has left and right attributes, where left and right are defined as they would appear to an observer sitting at a node (first point in the chain) and looking along the chain toward the last point in the chain. The attributes in each record are polygon descriptors, polygon numbers and polygon areas (in hectares). The number of points in the chain is also recorded in each Chain Data Record.

Given the general format standard for geocoded polygon data defined above, several decisions had to be made concerning specific format details during the demonstration of data transfer feasibility. To ease the requirement to handle any record lengths that are a multiple of 180 bytes, a 360 byte record length was adopted. The polygon descriptor lengths were allowed to be a multiple of 4 bytes but 20-byte descriptors were commonly used. In order to reduce space requirements on the magnetic tape media the Chain Data Records were grouped into blocks of 10 records and the last block was zero filled when there were less than 10 records in the block. It was decided that the polygons in the region of interest were to be surrounded by a neat line which was to close all polygons in a topologically correct manner at the edge of the scene. The polygon outside the neat line was to be numbered zero and was to have a null descriptor. Further comments on the impact of these decisions on future revisions of the defining document are given in the concluding section of this paper.

IMPLEMENTATION OF INTERCONVERSION SOFTWARE AT CCRS

Software to convert a thematic raster image to SDTC polygon format and back to grid form was implemented on the PDP11/70 computer of the CCRS Image Analysis System (CIAS) (Goodenough, 1979) in March 1981 by Collins and Moon Ltd. (CML) of Guelph, Ontario. Although internal (loop) tests were successful, numerous problems were encountered in rasterizing data received at CCRS from the CGIS and in reformatting data from CCRS into the CGIS data base format at Lands Directorate. Many of the problems were corrected by subsequent work funded by CCRS and performed by CCRS and CML staff. By August 1982 SDTC Format data from CGIS could be successfully rasterized at CCRS. At this time also, SDTC Format data from CCRS could be read at Lands Directorate and plotted on their Automap System. However, the data could not be used to produce thematic maps or be overlayed with other data after being entered into the CGIS data base due to a remaining topological problem. The following descriptions outline the current implementation of the interconversion software at CCRS. Figure 2 presents an overview of the interconversion software system.

Chain to Raster Data Conversion

In the rasterization process, the chain data records are first read and converted from UTM coordinates to corresponding line and pixel coordinates. The intervening points making up a polygon boundary are interpolated to line and pixel coordinates. The left and right polygon classifications at each boundary are at the same time associated with each boundary point. The points are then sorted into line and pixel order. From top to bottom of the image, and left to right on a line, pixels are assigned a number (which denotes a particular colour on the raster display) which depends on the active class as determined at each polygon boundary crossing. All configurations of polygon boundary crossings must be accounted for to ensure that no stripes are produced across the image. When attaching high resolution polygon data (10 metre say) to a lower resolution grid (usually 50 metre), several polygon boundaries may pass through a resolution element. Appropriate data structures and logic must be implemented so that the pixel to the right of the element in question has the class determined by the rightmost polygon boundary. The rasterization algorithm also requires that all polygons be closed; i.e. that all of the chains making up a polygon are accouuted for and that the end points of the chains coincide at nodes.

Raster to Chain Data Conversion

The raster to chain data conversion program scans the raster image line by line with two lines of data being retained in memory at all times. Each pixel is compared to its northern and eastern neighbours. If the pixels belong to different classes, a polygon boundary is recorded between the differing pixels. The polygon boundaries are 'strung' in this way with new polygon numbers being assigned when needed as the image is scanned line by line. A newly created polygon may, further down the image, be found to be the same as a previously defined polygon. These are identified and later made equivalent by defining a new polygon having the same total area. Finally the polygon coordinate strings are sorted and 'coalesced' into chains. These chains should include all the points defining the common boundary between two polygons from one node to another.

There are three other programs used to process Chain Data Files and SDTC Format tapes. One of these reads SDTC Format tapes and places the Chain Data Records on disk in a Chain Data File. It also writes Chain Data Files to tape and constructs the Volume Descriptor File and the File Descriptor Records. The working Chain Data File on disk has 4-byte binary integer polygon descriptors in its Chain Data Records to ease data handling and computational complexity. However, output tapes are produced with 20-byte ASCII descriptors and input tapes can have ASCII descriptors which are any multiple of 4 bytes (up to 40 bytes). It is possible to convert between files having differing descriptor lengths and different types of encoding. Finally, a plotting program is available to produce electrostatic plotter output of Chain Data Files by drawing polygon boundaries at any desired scale.

The programs described above underwent extensive development and redevelopment phases. Some problems encountered which are not directly related to the format standard may be of interest to those

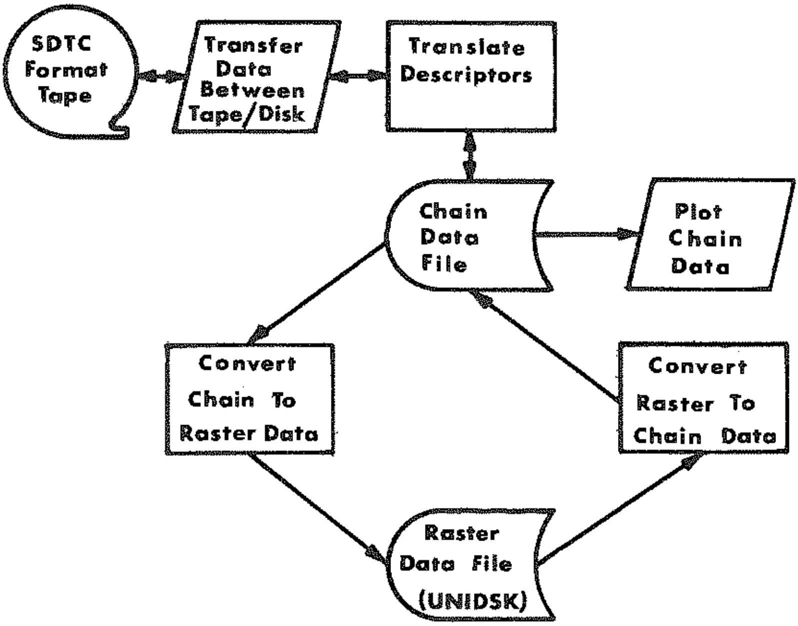


Figure 2. SDTC Format Interconversion
Software Overview

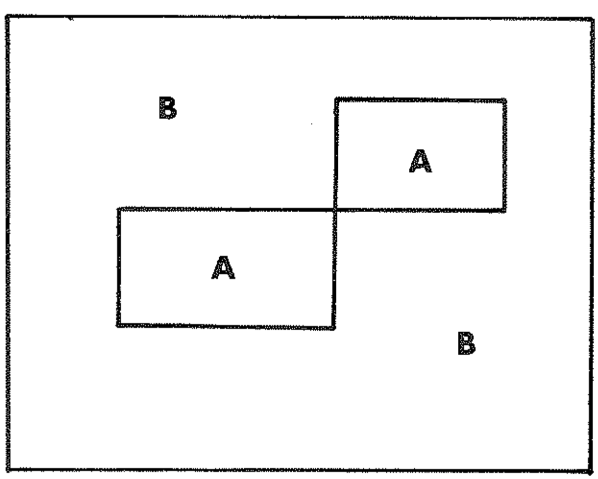


Figure 3. The 'Figure Eight' Problem

contemplating work in this area. One difficulty in rasterizing polygon data when the input chain data is at a higher resolution than the raster grid is the disappearance or breakup of aubpixel width features such as rivers. Algorithms were implemented to allow for the prevention of 'collapse' of specified features, for example, constraining any pixel through which a river boundary passes to be assigned to the colour code for water.

Another difficulty was encountered when duplicate chains were found in data from Lands Directorate, in that the requirement noted above for proper rasterization, that all polygon boundaries must be closed, would be violated. This problem is one which is difficult to detect because duplicate points are not evident on plots of chain data. No duplicate chains were found in data received from Lands Directorate after July 1982.

Some of the major problems in raster to polygon conversion were in data handling. In a complex classified image, many polygons encompassing only a few pixels can be present. As a result, for a 1000 by 1000 pixel image the Chain Data File can occupy several times as much disk space as the corresponding raster image. The data was compressed somewhat by removing all points lying on straight lines except the start and end points. Generalization of classification results was found necessary in most cases in order to process images in sensible execution times and using reasonable amounts of disk space. Median filtering was employed for this generalization; i.e. each pixel value was replaced by the median of the pixel values in its neighbourhood. Although this type of filter eliminates small isolated groups of pixels, classification boundaries are generally altered. More acceptable solutions would be classification filtering with a loss matrix operating on the raster image (Goldberg and Goodenough, 1976) or a small-area deletion algorithm operating on the Chain Data File.

The remaining topological problem in the SDTC Format data presently produced at CCRS which prevents the data from being used in the CGIS is due to a particular design premise in the original CML software. Their design strove to maintain the selectability of linear features such as rivers by a single polygon number. This can be referred to as the 'figure eight' problem when the resulting topology is compared to that required by the SDTC Format.

As shown in Figure 3, in the original implementation, the two rectangles forming a figure eight and labelled class A would be considered to be one polygon with one associated polygon number and with all 9 points defining it in one Chain Data Record. However, according to the SDTC Format standard these are two separate polygons which should be written into two Chain Data Records since the rectangles are touching at a node. In February 1983 the raster to polygon conversion software was modified so that the two rectangles would be treated as isolated polygons. However, after doing this the point of contact can not then be identified as a node and the chain data records in many cases are not written out with the node as the first point in the record. In order to conform to the standard which specifies that the first and last points of each chain must be nodes, an extensive revision of the software would be necessary. Alternatively, for efficiency, one could make a revision of the standard to orient it more explicitly to polygons. The STDC must make a decision on this before the next revision of the format standard.

The topological problem described above does not prevent the simpler process of the production of thematic maps by the GIMMS system (Waugh, 1982) at Statistics Canada however, and the successful transfer of data to that system is described in a companion paper (Korporal, 1983). That paper also shows a thematic map and a plot of its associated polygon boundaries produced from SDTC Format data received from CCRS. Unlike the CGIS, the GIMMS system does not produce raster thematic maps from a segment data base but generates an intermediate file of closed polygons which is then displayed on a colour terminal. The topological errors do not interfere with the generation of these polygons.

CONCLUSIONS

As noted in the previous section, a redesign of the CIAS resident software would be necessary to satisfy all of the requirements of the SDTC Format standard. However, there are several limitations of the PDP11 computer in this application which have lead to a decision not to carry out the revision on that system. The major sources of difficulty is the limited (64K) address space of the PDP11. As a result, data which ideally should be held in memory during processing must be kept in temporary work files. In effect, a virtual memory system had to be implemented to accomplish this processing task. A similar 'virtualization' technique was implemented for a raster to polygon conversion software system by Nichols (1982).

The overhead for paging of memory in turn increases processing time. For a 1000 x 1000 pixel image consisting of about 10,000 polygons the processing time is about 10 hours and the temporary disk files can be up to 50,000 blocks (25 Megabytes) in size. Hoping to overcome some of these problems, we are presently designing software and procedures to produce SDTC Format data by postprocessing of the polygon data from the map input-output subsystem which is part of the LANDSAT-4 Digital Image Analysis System (LDIAS) now under development at CCRS. This VAX11/780 system should be effective in this application due to its greater computing power and available address space.

Finally, there are several revisions to the SDTC Format document and subsequently to the design of the new software that have been found necessary during the data transfer effort. The suggested revisions are in the following areas:

1) Blocking of logical records into the same physical record is not part of the superstructure format. Changes need to be made in field definitions to allow incorporation of this feature under superstructure restrictions.

2) Some information currently stored in the Master Header Record would be more appropriate in the Geographic Detail File Descriptor Record to be more in accordance with the superstructure standard.

3) The superstructure format also demands that the first 12 bytes of every record be reserved for record number, record type codes and record length information. This convention is not presently followed for the Chain Data Records.

It is hoped that a future revision of the SDTC Format document will incorporate these changes as well as other minor corrections.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the contributions of Kate Dickinson to the work described here.

We thank Lynn MacFarlane and Mike Robson for their assistance with the data transfer effort. We also thank Mike Comeau of Lands Directorate,

Environment Canada for his cooperation and for his work in identifying problems with our data during several demonstration transfer attempts.

REFERENCES

Goldberg, M. and D.G. Goodenough, 1976, Analysis of a Spatial Filter for LANDSAT Imagery, SPSE Conference Proc., p. 276-282, Toronto.

Goodenough, D.G., 1979, The Image Analysis System (CIAS) at the Canada Centre for Remote Sensing, Canadian Journal of Remote Sensing, Vol. 5, No. 1, p. 3.

Goodenough, D.G., K.J. O'Neill, L.A. Gordon, J. Yan, T. Fisher, C.L. MacDonald and A. DesRochers, 1979, Standard Format for the Transfer of Geocoded Information in Spatial Data Polygon Files, CCRS Research Report 1979-3.

Korporal, K.D., 1983, A Demonstration Transfer of Remotely Sensed Data Utilizing the Standard Format for the Transfer of Geocoded Polygon Data, Proceeddings of Autocarto VI, Ottawa, Canada.

LGSOWG, Configuration Control Board, 1979, The Standard CCT Family of Tape Formats, Document Number CCB-CCT-002C.

Nichols, D.A., 1981, Conversion of Raster Coded Images to Polygonal Data Structures, Proceedings of Pecora VIII, p. 508, Sioux Falls, SD.

Waugh, T., 1982, GIMMS Reference Manual, GIMMS Ltd., Edinburgh, Scotland.

A DEMONSTRATION TRANSFER OF REMOTELY SENSED DATA UTILIZING THE STANDARD FORMAT FOR THE TRANSFER OF GEOCODED POLYGON DATA

Kenneth D. Korporal
Department of Geography, Carleton University
BP 254, RR#1
Cantley, Quebec J0X 1L0

ABSTRACT

As part of a graduate research program at Carleton University, a cooperative project to transfer remotely sensed data from the Canada Centre for Remote Sensing (CCRS), Department of Energy, Mines and Resources to a geographic information system at the Geocartographics Subdivision (GCG), Statistics Canada, was undertaken in early 1982. Utilizing ad hoc standards developed by a committee of federal government departments, the "Standard Format for the Transfer of Geocoded Information in Spatial Data Polygon Files", a series of increasingly complex files have been transferred and subsequently mapped. LANDSAT Multispectral Scanner (MSS) data was processed, classified and manipulated. Vectorization processes and special software to create the standard format transfer files were developed by CCRS (as described in a companion paper by Goodenough et al, 1983). Software interfaces were developed at GCG to facilitate use of the data in GIMMS, a comprehensive geographic information mapping system. Subsequent processing produced high-quality thematic maps and analytical graphics.

INTRODUCTION

The impetus for attempting a transfer of remotely sensed data was spawned during the course of a graduate research tutorial which was examining relationships between computer cartography and remote sensing as they existed in 1982 (Korporal and Manore, 1982). It was known that the Standard Data Transfer Committee (Goodenough et al, 1979) had developed a format which was accepted by all member agencies as well as other federal and provincial organizations. Although the Standard Data Transfer Format (SDTF) was accepted as a standard, no agency to our knowledge, had successfully completed a data exchange using the SDTF. During the course of this graduate tutorial program it was realized that remotely sensed data in the STDF could be used as an input into GIMMS (Waugh, 1982), a geographic information and mapping system.

GIMMS is a very large, user-friendly, integrated geographic information mapping and manipulation system. The program has been developed over a number of years and presently consists of a number of inter-connected applications including mapping, graphics, data manipulation, and interactive design components. The mapping and data manipulation applications are perhaps the best developed of the system and are those which have the most immediate relevance to the data transfer. Being device independant, GIMMS can produce medium to high quality output on a number of line drawing devices, utilizing black and white or colour shading patterns, point symbols, dot patterns, and/or solid colour if available.

In the context of the tutorial project, an experimental transfer was designed utilizing the facilities at the Canada Centre for Remote Sensing (CCRS), Department of Energy, Mines and Resources, and the Geocartographics Subdivision (GCG) of Statistics Candada. Both agencies were very supportive of the project, donating computer and other machine time, program support, and valuable assistance. Following the first set of experimental transfers, improvements were made both to the programs residing on the GCG Vax 11/780 and to those resident at CCRS, thus providing an opportunity to test new and larger data files. This paper will describe the first transfer attempts, subsequent developments and demonstrations, and examine planned and potential future enhancements for an automated transfer and geographic information/mapping system.

THE FIRST TRANSFERS

The first transfer attempts successfully achieved the objective of transferring remotely sensed data from one agency to another and subsequently inputting the data into a geographic information and mapping system. In order to meet the primary objective of mapping remotely sensed data within the strict time-frames imposed in the academic environment, program development had to be kept to a minimum. By taking advantage of utility programs resident on the DEC computers (PDP 11/40 at CCRS, VAX 11/780 at GCG) the first series of transfers required only minimal program development for data receipt and display. Program development was required to convert the data from mixed ASCII and BINARY to ASCII files formatted for input into our mapping system, GIMMS. Many problems were encountered at all phases of the process but most were rectified as they were discovered. The most serious problem existed in the data output from the raster to vector conversions at CCRS where it was found that topological errors were created which prevented formation of closed polygons as required for thematic map production. Nonetheless, the problems were resolved sufficiently to permit a simple thematic map to be produced and fulfill the primary course objective (see Figure 1). Subsequent developments and demonstrations have differed largely in one respect: the SDT format has been used instead of the DEC dependent formats of the first transfers.

STANDARD DATA TRANSFER FORMAT

The Standard Data Transfer Format (SDTF) uses polygon chains as the basic structure around which the format was developed. The SDTF also contains general geographic, spatial and descriptive information while conforming to standards developed for geocoded Computer Compatible Tape

(CCT). Design of the STDF incorporates several important properties essential for present and future applications including machine and language independence, generality, expandability, and, self-definition (Goodenough et al, 1979).

At present, spatial data in the SDTF is composed of three files; the Volume Directory File, the Master Header File and the Geographic Detail File, which includes the detailed polygon chain information. For demonstration transfer purposes this third file is the most important because in addition to the actual X and Y coordinates, this file contains the chain numbers, left and right polygon descriptors, left and right polygon numbers, number of points in the chain, and left and right polygon areas. This is the essential geocoded information required for input into GIMMS.

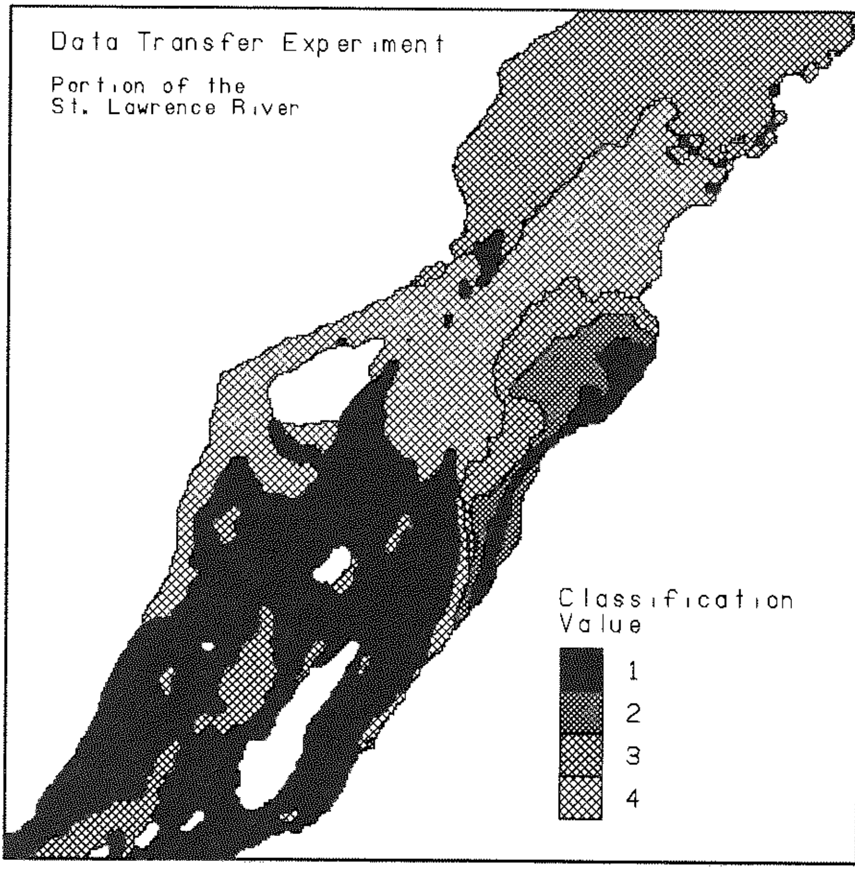


Figure 1. One of the first thematic maps produced from transfer data.

PROCESSING AT CCRS

The remotely sensed data used for SDTF transfer purposes requires minimal special processing other than the final formatting and transfer to tape. Although any classified digital data can be used, LANDSAT Multispectral Scanner (MSS) data was used for most of these demonstration transfers. Standard image processing and classification procedures are utilized by CCRS for creating a thematic raster image.

For demonstration transfer purposes, some theme manipulation was required to remove single or small groups of scattered pixels causing "speckle" in the image classifications. Two reasons for this are (a) cartographic generalization for thematic mapping, and (b) the need to avoid excessive data storage and computing requirements. Generalization of thematic data into contiguous classes may improve map readability and perhaps better represent real conditions. If, however, patterns produced by "spurious" pixels are important, then such theme manipulation may not be beneficial.

After completion of the theme manipulation and processing, the thematic data is vectorized using raster to polygon software implemented for SDTF purposes. The final step at CCRS is writing the vectorized data to tape using software which converts the data to the SDTF description. A more detailed outline of the processes and the SDTF software is given in a companion paper (Goodenough et al, 1983).

PROCESSING AT STATISTICS CANADA

Before processing of the remotely sensed data at Statistics Canada could be undertaken, some software development was required to read and format the data from the SDTF tape. A VAX 11/780 mini-computer is the base for the software, including the thematic mapping program and the new software written for this demonstration transfer.

The GCG processing can be divided into four major operations:

(a) reading of the SDTF tape
(b) converting and formatting the digital files
(c) line and polygon file creation
(d) thematic mapping and graphics

Figure 2 illustrates these major processing steps at Statistics Canada.

<u>SDTF Tape Reading</u>
Reading of the tape is a relatively straight forward matter using standard system VAX foreign tape mount procedures. Normally, the tape will have three files giving volume, header and segment information. Details essential for accurate reading such as record lengths, number of records, and some geographic information including projection, scale, etc. are on the first two files. A program called SDTREAD was developed to convert the files from mixed ASCII

and BINARY to ASCII files. The third file, Geographic Detail File, is processed separately because it relies on information contained in the first two files.

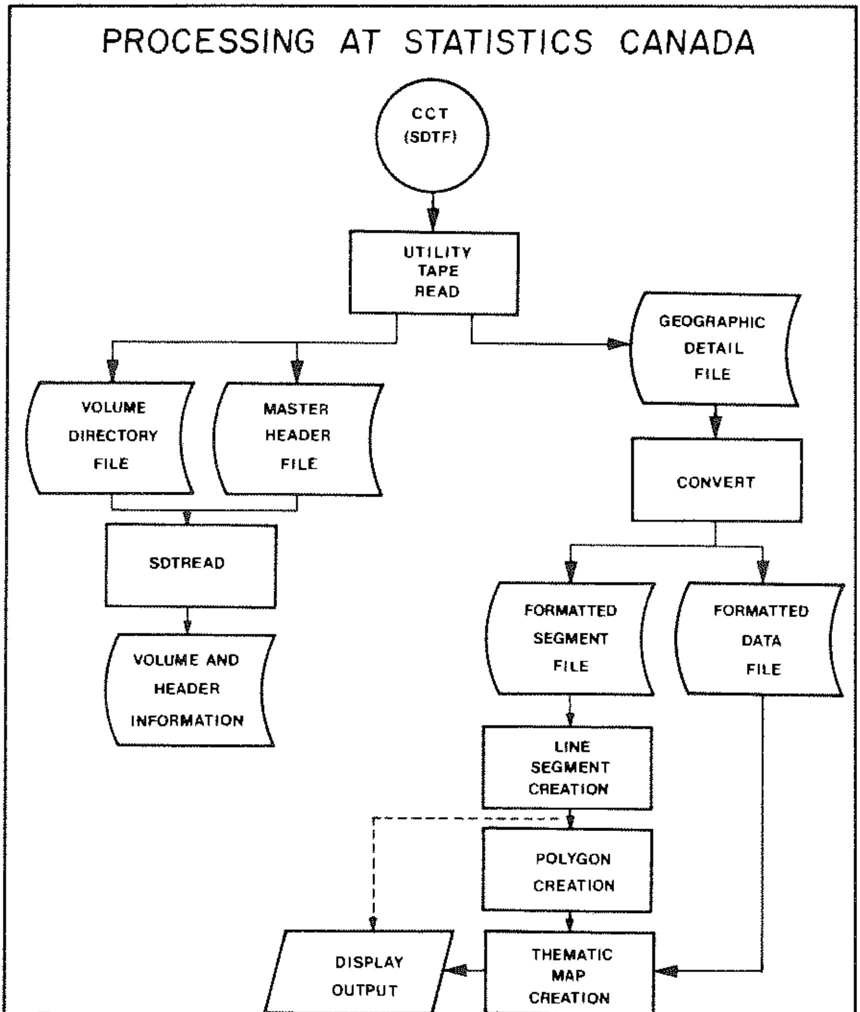


Figure 2.

File Conversion and Formatting

Reading, converting to ASCII, and formatting of the Geographic Detail File presented the most difficult and time consuming developments for the transfer. The program created to do this task is called CONVERT. Much of the difficulty in defining the program was due to the uncertainties of the quality of each step in the transfer process. For example, when spurious lines would appear in a file which was in the thematic mapping stage, it was difficult to determine if the problem was related to GIMMS, to our conversion program CONVERT, or to incorrect information on the SDTF tape data files. Therefore, considerable time was spent in tracing problems to the source.

CONVERT, in its present form, creates two output files. The first is a formatted segment file for GIMMS input with segment labels, coordinate pairs and unique GIMMS command identifiers. The second output file is a formatted and sorted data file of classification values and polygon labels, also in GIMMS input format specifications. The first of the above files is used to create a line segment file then a closed polygon file, while the second file is used only at the thematic mapping phase.

Line and Polygon File Creation

Upon completion of CONVERT processing, the segment data file is ready for GIMMS processing to create two more output files and, later, merge with the data classification file to produce thematic maps.

Using the CONVERT generated segment data file, a segment line file is derived which, when drawn (see Figure 3), maps the segments as encountered in the raw data file. In the context of thematic mapping, the main purpose of this file is as the input for creation of a polygon descriptor file.

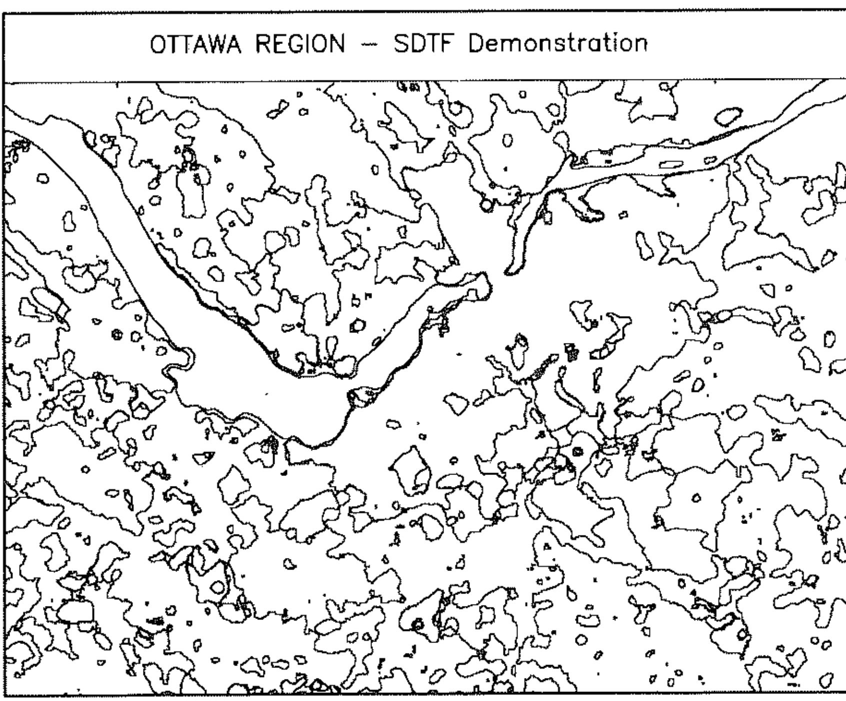


Figure 3. Plot of a GIMMS segment file - scale 1:250000

Polygon or area type files differ from the segment file in that a complete 'closed' boundary description is required where the start and end points for each polygon are identical. The polygon file is required for most thematic mapping purposes where shading or dot patterns are wanted. The very large size of the data files generated from remotely sensed data, combined with the many small areas described by small groups of pixels, puts great demands on GIMMS whose usual mapping application is for much larger areas such as those of political or census type boundaries. GIMMS has the capability to allow the easy upgrading of "normal" program limits for unusual or unique applications such as those encountered in this demonstration transfer.

An important feature of GIMMS is that a check can be made on the segment data to test for errors such as polygon closure, at much lower processing costs than required for creating the polygon data files. This check is recommended if errors are anticipated in the segment file description. In addition, a generalization function can be introduced into the polygon file creation process to reduce the number of points and generalize the boundaries, thus reducing the processing costs associated with the thematic mapping component.

Thematic Mapping

The GIMMS mapping module is capable of producing medium to high quality maps with supplemental graphics at virtually any scale or size, limited mainly by the capabilities of the output device. Cartographic enhancements are an integral part of the program enabling the addition of titles and other textual features, legends, north arrows, shielding, overlays, offsets, many text fonts, multi-map sheets, etc. User specified black and white or colour shading and dot patterns are used for the effective presentation of data.

The classification values contained in the classified data file, the second file output from CONVERT, are required at the thematic mapping phase. This is the thematic data which will be represented by the selected symbolism and on which the graphics and user specified manipulations are based. An example of a GIMMS thematic map produced from classified digital data is illustrated in Figure 4.

FUTURE DIRECTIONS

As a demonstration system, the developments at GCG are adequate and suitable for the intended purpose: to demonstrate the transferability of data using the SDTF description. A number of system enhancements have been identified and are presently in various stages of development.

System Automation

An automated production system is being designed to permit the transfer of any data in SDTF description for use in GIMMS. The system will be "fronted" by a user-friendly module which will guide the user through the relevant steps: tape read, SDTREAD, CONVERT, segment creation and polygon creation. However, the user will still require a knowledge of GIMMS to draw the polygon files or to produce thematic maps.

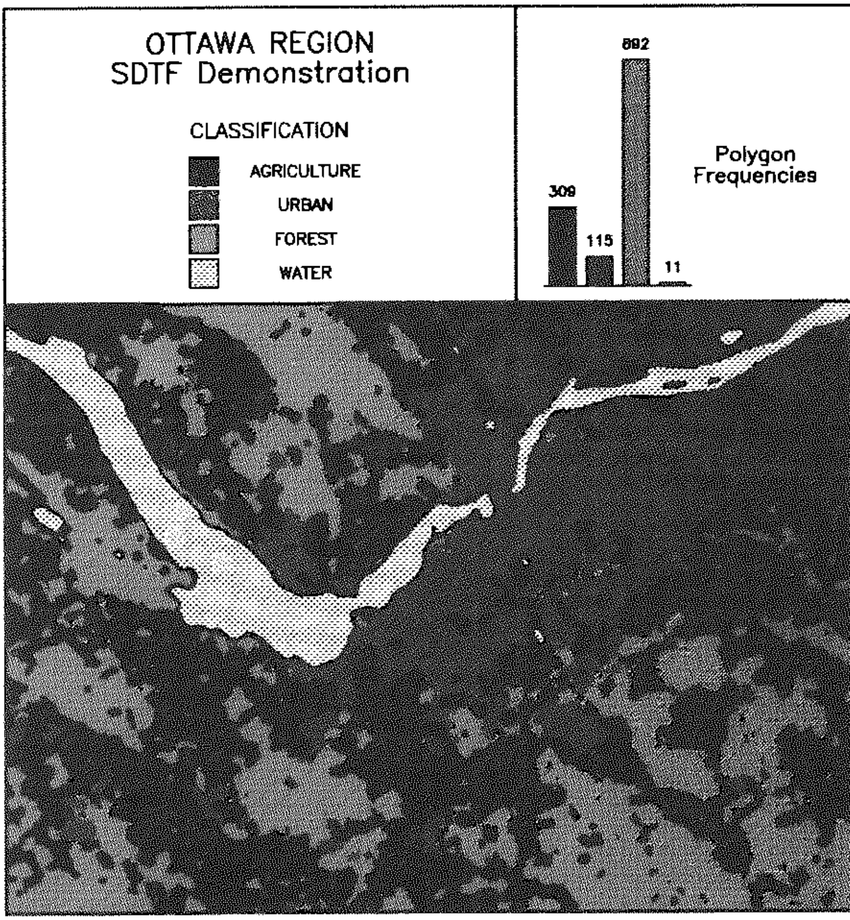


Figure 4. GIMMS thematic map and frequency histogram produced from CONVERT formatted remotely sensed data in SDTF.

<u>Two-Way Transfers</u>
To be a truly useful system there is a requirement for two-way data transfers. A number of steps are necessary, including:

(1) conversion of polygon data to segment data
(2) conversion of segment data to SDTF description
(3) conversion of SDTF data to raster image

Each of these components is available in either prototype or production systems in the two departments. PLUSX (Goodchild et al, 1982) can convert polygon data to a segment data format. Features within GIMMS provide the basis for conversion of segment data to SDTF based files. Programs have already been developed at CCRS for reading SDTF files and converting them to raster format. This task, therefore, is one mainly of moulding various independent programs into a modified format to create an automated system.

CONCLUSION

The transfer and mapping of digital remotely sensed data has demonstrated that the Standard Data Transfer Format is a viable working format. The very large file sizes inherent in digital data, such as classified digital images, provide severe testing of system abilities and limitations. While much development is required to create an automated production type system, this demonstration has shown that the advances to date are significant and that upgrading and improving on these developments could be implemented at other agencies. With the ability to quickly and easily exchange data between CCRS and Statistics Canada, and others adopting the SDTF, practical applications will become feasible without the acquisition of expensive and unique hardware.

ACKNOWLEDGMENTS

The author wishes to acknowledge the important contribution of Michael J. Manore as initial collaborator involved in the conception and the first set of experimental transfers. Special thanks to Dr. Gordon F. Deecker, Statistics Canada, for his programming expertise and valuable assistance. Thanks also to C. Shadbolt, M. Lemay and L. Brunet for cartographic and compilation assistance. And finally, to the many persons at CCRS who have been involved, including Dr. D.G. Goodenough, J. Palimaka, G. Plunkett and K. Dickinson.

REFERENCES

Goodchild, M.F., 1982, PLUSX Documentation: Department of Geography, University of Western Ontario

Goodenough, D.G., K.J. O'Neill, L.A. Gordon, J. Yan, T. Fisher, C.L. MacDonald, and A. DesRochers, 1979, Standard Format For The Transfer of Geocoded Information in Spatial Data Polygon Files: CCRS Research Report, Ottawa

Goodenough, D.G., G.W. Plunkett, and J.J. Palimaka, 1983, On The Transfer of Remote Sensing Classifications into Polygon Geocoded Data Bases in Canada: Auto-Carto VI Proceedings, Ottawa

Korporal, K.D. and Manore, M.J., 1982, An Experimental Transfer of Remotely Sensed Data: unpublished graduate tutorial, Department of Geography, Carleton University, Ottawa

Waugh, T., 1982, GIMMS Reference Manual: GIMMS Ltd., Edinburgh, Scotland

CARTOGRAPHIC FEATURE CODING

Bruce D. Wood, and David H. Douglas
University of Ottawa
Ottawa, Ontario, K1N-6N5

ABSTRACT

This is a short report on a research project which has investigated the problems associated with the development of a uniform topographic feature classification. Perspectives on the nature, issues and need for an integrated approach to feature classification are discussed. Five existing topographic feature classifications were critically analysed as part of the project and a series of nodule principles as well as some ideas on future development were generated to serve as a guideline for the implementation of a code which could facilitate the needs associated with both conventional map production and geographic analysis.

INTRODUCTION

The Federal Government of Canada, and other local and provincial levels of government, and industry, are in the process of encoding spatial data in digital form. A classification of topographic features has recently been proposed by the Canadian Council of Surveying and Mapping as a national standard. This standard, if adopted, will basically become the interchange linkage for all of the digital data collected by the Topographic Survey of the Surveys and Mapping Branch. Topographic Survey is, and has been in the process of digitizing terrain information since the 1960's. One of its present goals is to digitize map data of the whole country at the scale of 1:50,000. An investigation of this classification is, therefore, extremely timely. Governments within Canada, of course, are not the only ones at, or through, the threshold of providing map data in digital form. Indeed the whole problem of exchange of spatial information in digital form between governments and between the private and public sectors is looming monstorously on the information processing horizon.

The problems of spatial information technology exchange can be classified into four levels, elucidated by Witiuk (1976) as the conceptual level, the algorithmic level, the systems level and at the final results level. Seen in this context as occupying the lowest of the four may be misleading, in that, especially with spatial data, there are a host of conceptual problems with data which are easy to underestimate, and easy to discount when others complain of having them. This may explain the lack of attention given to data exchange. At a very basic level there are a host of technical standards regarding EDP mediums, formats, self defining headers and codes that must be developed and adopted, but these problems, although of concern, can be considered to lie largely outside the realm of cartography and within the fields of electronic data processing. Within the scope of cartography, the main concerns are those regarding data structure and the encoding methods for feature representation, and whether the features recorded are fundamentally the geographic phenomena themselves, or recordings of the traditional graphics utilized on paper maps to represent the geography.

HISTORICAL BACKGROUND

The problems of exchanging spatial data have come to the forefront of spatial information processing only recently, however in the broadest sense there has been a recognized need and an increasing attempt to exchange spatial data for over twenty years: Sparks, Guttenberg, Anderson, Clawson and Stewart to name but a few. These early commentators from the planning profession suggested that an important step in the right direction would be to develop a uniform land use classification. Two major handicaps they identified were: 1. the incompatible classifications and differing definitions of terms; and 2. the lack of measurement (Sparks, 1958, p.175). Similarly Guttenberg (1959, p.143) noted that since further progress in planning and planning research depends on exchange of empirical findings, it is especially important for planners to have a common language, a language so precise as to leave little room for misunderstanding. Anderson noted that the basic problem in the entire land use field is one of developing a classification with definitive criteria for separating the various classes that will give objective and repeatable results for any area being studied. A major work by Clawson and Stewart (1965) found that accurate, meaningful, and current data on land use are essential if public agencies and private organizations are to know what is happening and are to make sound plans for their own future action. At this early date they proposed that to overcome comparability problems over time and space, data on land could be transferred electronically to another point of use very quickly. "Thus it would no longer be necessary that each point of use have its own storehouse of land data in the form of tabulations, publications, maps etc., but rather, data obtained in one place or by one agency could be stored ready for use by anyone quickly and inexpensively." Is this not the basis of spatial data transfer in the digital era?

The problems associated with, and the importance of improving our ability for spatial data transfer are today accentuated by the fact that more and more government agencies, universities and private industry are making use of the computer for storage, retrieval and analysis of spatial data (Linders). Often however much data is collected, massaged, and stored without careful thought as to what other uses the data might be put. Several diseconomies associated with this phenomena include the simple (but expensive) fact that 1. often spatial data are not readily interchangeable (Yan et al.); 2. resources and accuracy are lost in the duplication of data collection and/or the extended efforts required to transform or regenerate data in order to use it; and 3. any real benefits derived from using the computer (such as timeliness, efficiency and/or replicability) are lost to the costs associated with outdated, incomplete and inaccessible data (Interministerial Committee on Geographical Referencing 1978).

Problems of exchange are increasingly being met by governmental committees established for the purpose. One of the earliest experiments in Canada was the formation of the Inter-Agency Spatial Data Transfer Committee which grew out of the less formal National Capital Geographic Information Processing Group of 1975. The Alberta Government, wishing to avoid the chaos of multiple development of incompatible systems endured by earlier entrants to the field set up the Land Related Information Processing Systems Coordination Project (LRIS) in 1978. Similar problems of exchange were recognized in Ontario, where, of sixty-four geographical referencing systems in the Government ... "it would be difficult if not impossible to transfer data between these systems without major modifications to most of them" (Interministerial Committee). One objective of this Committee on Geographical Referencing was to develop standards and specifications which could be incorporated into municipal, provincial and federal information systems in order to expedite the transfer and correlation of geographically referenced data with one of the prime concerns being the establishment of an acceptable grid reference.

Similar concerns expressed by the Report of the (Canadian) Task Force on National Surveys and Mapping revealed that a basic need at this time is the design and adoption of criteria for the storage, accessibility and exchange of digital spatial data among federal and provincial departments and agencies, utilities and the private sector. The Task Force noted that the most pressing need is the formulation and adoptation of standards for the communication of data. Of course other countries are also facing the same standardization issues and problems in similar ways. The ACSM National Committee for Digital Cartographic Data Standards is a case in point. All of these attempts to solve the data transfer problem by committees illustrate the importance of the subject, and indicate a measure of the intractability of the problems being faced.

THE FEATURE CODING PROBLEM

There are many aspects to the handling and transfer of spatial data. This short note attempts only to provide some definitive ideas on how the feature coding problem might be resolved. The digital description of a cartographic or geographic feature can be considered as having two components, a geometric description and an attribute description. The geometric description has been given rather more attention. Methods for numerically recording points, lines and surfaces have been developed to a high degree of sophistication. Topological structuring of data to represent networks, graphs, solids, etc., to enable more efficient searches and manipulations, to compute properties from the geometry and to generate interesting graphics are also well represented in the literature. Geometric description problems seem to offer more interesting challenges. This may be because of the multi-dimensional aspects of geometric data, but also because when a technician or researcher is working with a data-set it is usually of one kind of feature. It is a file of contours, or a digital elevation model of temperature, etc., and this fact reduces the importance of its identification since its description will be understood or may be noted with the written description of the file. Of course the importance of self describing data grows when one considers mass exchanges amongst governments and between the private and public sectors.

Problems of devising self describing codes for objects and their non-spatial attributes at first seems straightforward and perhaps rather obvious. If road maps are composed of subsets from fifty-seven different symbols, simply create a table of fifty-seven items, number them and include an integer from 1 to 57 with each geometric feature in the digital file. Sort them alphabetically before numbering them if it is expected that the list will be examined frequently by eye. Even the addition of that 58th feature need not cause excessive concern. Add it on the end, or assign to it a decimal number to insert it into its proper alphabetical place, or redraft the code attaching a flag to the data identifying the specific code list to employ. In fact, do not even use a numerical code, but rather let the alphabetic words themselves or mnemonic short forms act as the code. All of these solutions are indeed employed, but it is apparent that even at this level the problem is losing its trivial innocence. Contrasting with a road map series, a topographic series may have from 1000 to 2000 different types of features. The mere compilation of the list of things appearing on a whole series for a country the size of Canada is no trivial task.

As in traditional cartography, the topographic map forms the basic building block for thematic and special purpose maps. One contention is that the general purpose topographic feature classification adopted will have major implications on the level of geospatial analysis permitted with the encoded data in the future. This contention stems from three facts. Firstly, automated cartography is not limited to the automated drafting of maps but extends into the area of cartographic data services from which a wide variety of graphic products, both stored cartographic information as well as derived information, can be produced (displayed) on demand. Secondly, the topographic map, either directly or after scale and projection changes, provides the foundation for many of the published maps used increasingly by the many sciences and professions who demand terrain information (Zarzycki, Harris and Linders, pp. 1-9). Thirdly, any feature coding standard, if it is done well, may provide the foundation of other systems created for and by specific users, whether or not they utilize data from a Governmental Survey. This may create opportunities for transfers from non-governmental agencies, such as utility organizations, to the government for generalization to Survey needs.

Linders (1978, pp.188-191) notes that the development of a system for mass exchanges implies a taxonomy or classification of information with the objective of providing a single logical system for the storage and management of all land mass data. He suggested that such a system must encorporate detailed information for: feature location; feature taxonomy (the component elements of the data base must be uniquely classified through a data dictionary); attribute data (for further differentiation of features); relational data; and representational information (ie. the exact depiction of each feature element in terms of its graphic components).

Taxonomy implies the ordering of information into a hierarchy to develop understandings of phenomena, but in the context of cartographic or geographic features, to allow groupings on the basis of the criteria used to form the hierarchy. A tree structured ordering, for instance, could be made to allow a multitude of alias names or ordered groupings to result in the same classification. A digitizer operator may identify a feature with the words "thermonuclear power station," while another might say "atomic electricity plant." Logical algorithms could be constructed to automatically recognize both as being the same thing placing them into: "building: industrial, electric power generation station, thermal, nuclear". The coding system should be such to allow such algorithms to be written. On the retrieval side, like objects should be sequentially groupable on the basis of their identification or their properties. Beyond being possible to call all "jails and penitentiaries," it should be possible to call all "two to six lane roads," obviating the need to enter "three", "four", and "five". Furthermore, groupings across properties should be possible, such as a call for all "abandoned" things. These capabilities would be prerequisites for a system that would automatically filter and generalize data to prepare a file for maps at different scales. "Extract all information on transportation suitable for a map at the 1:4 million scale."

TAXONOMY AND CLASSIFICATION

Through the illustration of classification principles, one transends various realms of knowledge. Dolby (1979, pp.167-193) for example distinguishes the logico-mathematical theory of classification of the sciences, (for example the classification of the objects of study within a particular science as in the biological classification of living organisms), from related practical classifications intended for particular purposes, such as library science.

Taxonomy basically deals with the classification of all living things according to observed natural, or hypothetical, relationships, or both. The idea of taxonomy was first made explicit in the history of Western thought by Aristotle in his ORGANON and METAPHYSICS. The ruling principle is that the highest genus is divided by means of differentiae into subaltern genera, and each of these is then divided and subdivided until the ultimate species is reached. This principle has been handed down through the Stoics, Porphyry and the Greek commentators to Linnaeus, from whom it passed into modern biological usage (Peck, 1965, pp. v-viii).

Grigg notes that classification, defined as the grouping of objects (elements) into classes (sets) on the basis of common properties or relations, is a necessary preliminary in most sciences, and it is often argued that the state of classification is a measure of the maturity of a science. Although classifications can be built on various principles (morphologic, generic, temporal, spatial, quantitative, etc.), all must follow certain general and unalterable laws of logic: 1. the sum of classes must be equal to the scope of the classified generic concept; 2. only one classificatory criterion should be used within any one level of classification; and 3. a classification must not skip logical levels (Armand, 1965, p.22).

The logico-mathematical theory of classification, coincides with what, in the mathematical theory of sets, is called a partition. A division of a set of objects into subsets is a partition if and only if: 1. no two subsets have any elements in common; and 2. all of the sets together contain all of the members of the partitioned set. To the rules that a successful classification must be mutually inclusive and collectively exhaustive, Wynar (1980, pp.400-402) adds that notation must be flexible if the classification scheme is to be current. The classification scheme must also employ terminology that is clear and descriptive, with consistent meaning for both the user and classifier. Although most traditional classification schemes are based on a logical division of the universe of knowledge; by contrast Wynar states, computer-based classification systems are empirical and descriptive, attempting to develop thesauri with but one thing in common: a set of descriptors well suited to manipulation.

The following list of ten principles that should be taken into consideration when designing a classification system was proposed by Grigg (1965, p.481).

1. classifications should be designed for a specific purpose;

2. objects which differ in kind will not easily fit into the same classification;

3. classifications must be changed as more knowledge is gained about the objects under study;

4. the differentiating characteristics should be properties of the objects classed;

5. in logical division the division should be exhaustive;

6. in logical division and classification the species or classes should exclude each other;

7. in division, the division should proceed as far as possible upon one principle;

8. the principle of division must be important for the purpose of the classification;

9. properties which are used to divide or classify in the higher categories must be more important than those used in the lower categories;

10. the logical consistency of the hierarchy will only be maintained if rules five through nine are observed.

Classification obviously is very complex, based on logical principles which necessitate much conceptual foresight in both design and implementation. Problems inherent in any classfication system vary in extreme depending on the initial objectives of that science or discipline. These problems are more manifest in a general purpose classification because of the attempt for broader applications.

CARTOGRAPHIC FEATURE CLASSIFICATION

The following is a very brief summary of a systematic analysis of several significant cartographic feature codes now in use, or in the proposal stage. The objective was to identify an optimal code that was inclusive, flexible and open-ended, that could possibly be used in a general system, or at least to identify principles towards the development of one. Review, analysis and evaluation of these systems placed emphasis on identifying overall strengths as well as pin pointing weaknesses and faults.

The U. S. Geological Survey Attribute Codes for Digital Line Graphs (1980) consists of approximately 450 7-digit real number codes arranged into 11 base categories by the point, line or area characteristic of the feature classified. The implementation of parametric, multiple feature and coincidence coding, the integration with other codes (such as FIPS, Public Land Use and State Plane coordinate Zones), and the use of nodes as feature types to identify changes in feature classification are important assets. The main disadvantage relates to the geographic description of the feature being buried in the feature code itself.

The DMA Catalog (1977) shows us that it is possible to integrate a feature classification to facilitate the needs of three mapping centres. The schema is a hierarchy of 10 categories, 34 sub-categories, 97 classes and 753 features (each of which is identified by a 4-digit integer code) as well as the extensive use of attribute lists (3-digit integer codes). Its main advantage is the flexibility in feature description afforded by the attribute lists. Its main disadvantages relate to the inconsistency in feature description (some features are tersely described, others are but lengthy conglomerations of modifying concepts), and the double coding of similar attribute values with different 3-digit codes. The concept of

4-digit integer values, because of its limitation of 10 items per any digit will have to be modified to allow expansion in future versions of the standard.

The Australian Standard (1981), consisting of 657 4-digit feature codes, is quite similar to the DMA system but much simplier in terms of information provided. It lacks the comprehensiveness of the DMA catalog, particularly as this relates to feature modifier 'lists'. The fact that feature modifiers must be user defined offers some flexibility, but the fact that they are not standardized reduces their importance as far as a national data base is concerned. A similar version of the standard, provided by Systemhouse Ltd. of Ottawa, is in place in India. The concept of table driven parameters as well as certain relational aspects of the code merit investigation.

The Ordnance Survey Code (Digital Data Supplied to their Customers) is unique for its apparent simplicity. Quite likely the sophisticated aspects are transparent to the user. The Ordnance Survey supplies the graphic program with their digital topographic data. Parent scale series are 1:1250, 1:2500 and 1:10,000 for urban, rural and wilderness mapping respectively.

The Canadian Hydrographic Service Feature Classification, designed purely for paper chart production via a digital data base, provides some good ground for comparison with the Canadian topographic feature classification, even though the two classifications are incompatible. C.H.S. store their attribute data in a file separate from the geometric data.

The Canadian Topographic Feature Classification consists of 1558 10-digit alphanumeric codes, hierarchically arranged according to the levels of class, category, feature and attribute. As a first draft for a national standard, it has made an exemplary and fundamental contribution by compiling all items which occur on the traditional paper graphic for the whole series of Canadian topographic maps. As a national standard however, notable weaknesses which require re-examination include: 1. reduction of data measurement level, sometimes arbitrarily dictated by alphabetical arrangement; 2. fixed limitations regarding expansion of new types of features (either the features fit with the structure as proposed or they won't be able to fit); 3. unreferenced redevelopment of classifications highly refined by other agencies, such as the SIC code; 4. fixed limitations, by its structural design, concerning relations (combinations) of geographic descriptors; 5. a reliance on an arbitrary definition of what is defined as topographic and what is thematic, thereby artificially reducing the complexity of the problem; 6. the mere representation of features appearing on paper maps rather than a consideration of what is appropriate for digital map products; and 7. imbalances in the level of detail by which certain features are differentiated as compared to others.

SOME LESSONS LEARNED

A hierarchical arrangement of items, classes, categories, etc., of features and their attributes implies a tree data structure. Further, a requirement to be able to add and subtract attributes and features to any set of them implies list processing. Any numbering or coding scheme of this complexity should obviously be machine computed. Exact codes should be calculated by an algorithm, with the exact numbering system, and the numbers themselves, being of little concern to any user. The concern of the designer should be directed to the task of providing algorithmic consistency and not to the numbers or codes. The task of assigning numbers and working with them, in fact, becomes so onerous that shortcuts are invariably taken by those who go at it that way. For instance it becomes convenient to assign codes with fixed numbers of digits, with as few digits as possible, and as one number, all to capitalize on the eye's sensitivity to the graphic symmetry of a table, and to reduce clerical exhaustion. This leads to such shortcuts as assigning one digit to identify members of a set when there is reason to believe there will be less than eleven items in the set. When that eleventh element comes along all sorts of games are played to avoid the friction of adding that extra digit. Invariably attempts are made to fit it in as a member of a previous set, or the members of the previous sets are rearranged, or the code number gets replaced by an alphabetic, or alphanumeric number. This may expand capability to 26 or 36 elements before saturating the code, but all sorts of headaches are created for subsequent software development.

List processing by hand is tedius. Insertion of an item implies retyping the list. Items following the insert have to be renumbered, and these changes are seen to conflict with the need for stablility of the code. The initial numbering of items with an extra zero on the lessor significant digit side allows for the insertion of ten items, but only if these ten items arrive in the order they are to be inserted. This is obviously not satisfactory.

All of these things point to the one great pitfall of precipitously becoming involved with assigning numbers or alphanumeric codes to create the system desired, because the problems of manipulating these codes soon consumes all attention and begins to shape the hierarchy or structure that is being constructed. We have the tail wagging the dog. All of these problems are false ones in the context of modern information processing techniques. Regarding efficiencies in the actual coding, we must be concerned with real, and not false efficiencies. Another byte is a small price to pay for just one more advantage, but if it means that to conserve that byte we must abandon a logical system, that effort is utterly pointless.

There may be several valid hierarchies within which a list of items may be organized. If the purpose of a list of cartographic symbol features is purely for the printing of paper maps a perfectly valid hierarchy could be based on the printing process. Items could be divided by the colour of plate, by intensity screens, by symbol type. On the other hand, a spatial set theory approach on the basis of map space could be utilized to construct a hierarchy that exhausts the total area of the map space. A "college" would then be included in the group defined by "designated area: Educational", which would fall in "built-up area", which would fall in "land area" (as opposed to "water area"). Several valid hierarchies could operate in parallel, and that implies that mechanisms to point relations between these hierarchies could be devised, perhaps expressed in the notation of Francois Bouille (1978) as a hypergraph based data structure. Most importantly, it can be seen that complete hierarchies can be built on a single set of criteria. Trouble arises (meaning severe limitations on the usefulness of the result) when classifying criteria are not explicitly and rigourously defined and attempts are made to create hierarchies on mixed criteria.

A number of cartographic features have attributes that are numerical. A contour has a height, a road has a number of lanes, a railway has a guage, etc. Other measures are less sophisticated, but nevertheless maintain order, eg. "Jail complexes: Federal, Provincial, County",... "Survey monuments: 1st order, 2nd order, 3rd order, Doppler". However most attributes are nominal, eg. "Mines: Copper, Gold, Iron, Silver, Uranium". Because most are nominal there is a temptation to treat everything as nominal, and this can lead to serious and unnecessary losses of information and manipulative capability for those of a higher order. An exaggerated example of unnecessary conversion from higher order to lower order would be to take a thermometer (based on an interval scale of measurement) and assign range nominal values such as FR (frigid), CO (cold), NI (nice), HO (hot), and SW (sweltering), and then to sort these alphabetically: CO-FR-HO-NI-SW. When the order is lost, ie. conversion from ordinal to nominal, ranging capabilities are eliminated. When the interval value is lost by classifying, other classifications are excluded. When ratios are lost by poor definition of zero, much capability for mathematical manipulation is foregone.

Consideration of ratio and interval numbers as attributes necessitates the use of a "unit of measure". An attribute does not only have a "value", but also a "value description". Often a proper "unit of measure" can answer description questions. "Contour: 100" becomes "Contour; fathoms: 100.0". Formal inclusion of an attribute description, in the manner of a "unit of measure", applies to nominal attributes as well. "Factory: Cement blocks", becomes "Factory; Product: Cement blocks" eliminating ambiguity over to what, exactly, "Cement blocks" refers. Items may also have multiple attributes utilizing an attribute description, eg. "Rail line; tracks: 3; guage(meters): 1.4351, status: Abandoned". Attributes can also have multiple values, eg. "Hall; Function: community, dance, exhibition". Multiple attribute and value conventions would eliminate the need for hundreds of special feature codes representing mere combinations, which, when combined with coding restrictions, may lead to the exclusion of important attribute and value combinations. In the proposed Canadian code, for instance, it is not possible to record a "proposed one-way road" because the other combinations of one-way, elevated, numbers of lanes etc., saturate the digits allocated for features.

CONCLUSIONS

This brief paper has attempted to: 1. provide some definitive ideas on how to solve the problem of adopting a general purpose topographic feature classification system that is inclusive, flexible and open-ended; 2. emphasize the relevance of the map coding problem to those disciplines concerned with the scientific description and explanation of spatial relations; 3. demonstrate that particular aspects of Canada's proposed national classification should be reconsidered; and 4. strengthen the theoretical framework underlying code development.

This research is important because the whole question of standardization in Cartography is important. Traditional cartographers have always been concerned with the standardization of map symbols. Standards for digital cartography require new perspectives which will require philosophical approaches to the problem of classification. A cursory examination of the problem reveals that some sort of continuing body such as afforded by national committees of individuals representing a spectrum of map making, map using, and cartographic research disciplines is necessary if some kind of cartographic feature classification system is to find acceptance and be maintained. It is also apparent that the problem is one that requires profound academic and scientific attention at this stage.

ACKNOWLEDGMENT

The authors would like to acknowldege Mossad Allam, Francois Bouille, Nicholas Chrisman, Robert Rugg, Lou Sebert, and Ewa Siekierska, who gave up some of their valuable time for interviews and for their ideas.

REFERENCES

Anderson, James R. 1961, Toward More Effective Methods of Obtaining Land Use Data in Geographic Research: Professional Geographer, Vol. 13, No. 6, pp. 15-18

Armand, D. L. 1965, The Logic of Geographic Classifications and Regionalization Schemes: Soviet Geography, Vol. 6, No. 9, pp. 20-38

Bouille, Francois. 1978, Hypergraphs and Cartographic Data Structures: The HBDS System - ICA 9th International Conference on Cartography, College Park Maryland, pp.1-18

Canadian Council on Surveying and Mapping. 1982, Standards for the Classification of Topographic Features - Draft Report of Technical Committee 1, Energy, Mines and Resources Canada, Ottawa

Canadian Hydrographic Service. 1983, Feature Code Manual: Cartographic Research, Fisheries and Oceans Canada, Ottawa

Clawson, Marion and Stewart, Charles. 1965, Land Use Information: A Critical Survey of U.S. Statistics Including Possibilities for Greater Uniformity, (2nd ed., 1969), Resources for the Future Inc., Johns Hopkins Press, Baltimore

Defence Mapping Agency. 1977, Standard Cartographic Feature Digital Identification Catalog, Washington, D.C., Defence Mapping Agency

Dolby, R. G. A. 1979, Classification of the Sciences - The Nineteenth Century Tradition, Roy F. Ellen and David Reason, eds., Classifications in Their Social Context, Academic Press, London, pp. 167-193

Guttenherg, Albert E. 1959, A Multiple Land Use Classification System: Journal of tbe American Institute of Planners, Vol. 25, No. 3, pp. 143-150

Inter-Agency Spatial Data Transfer Committee. 1979, Standard Format for the Transfer of Geocoded Information in Spatial Data Polygon Files: Energy, Mines and Resources, Document Code For-SDP-0001, Ottawa

Interministerial Committee on Geographical Referencing. 1978, Proceedings, Ontario Ministry of Natural Resources, Toronto

Grigg, David. 1965, The Logic of Regional Systems: Annals of the Association of American Geographers, Vol. 55, No. 3, pp. 465-491

Linders, James. 1978, Geo Referenced Data Bases: Proceedings of the Interministerial Committee on Geographical Referencing, Toronto, Ministry of Natural Resources, pp. 183-214

Ordnance Survey, Automated Cartography Unit. 1980, Ordnance Survey Digital Data - Feature Codes and Description: Appendix D, Digital Mapping Facilities For Customers, Fixed Price Services Section, Ordnance Survey, Romsey Road, Maybush, Southampton

Peck, A. L. 1965, Introduction, *Aristotle*: *Historia Animalium*, (translated by A. L. Peck), Vol. 1, Heinemann, London

Report of the Task Force on National Surveys and Mapping. 1978, A Study Conducted for the Department of Energy, Mines and Resources, Surveys and Mapping Branch, Ottawa, Contract No. OSQ77 - 00013

Sparks, Robert M. 1958, The Case for a Uniform Land-Use Classification: *Journal of the American Institute of Planners*, Vol. 24, No. 3, pp. 174-178

Standards Association of Australia. 1981, *Interchange of Feature Coded Digital Mapping Data*: Australian Standard AS 2482-1981, Standards Association of Australia

U.S. Geological Survey Topographic Division. 1980, *Computer Files and Attribute Codes for Digital Line Graphs*: U.S. Geological Survey Topographic Division, Reston, Va.

Witiuk, Sydnew W. 1976, Issues in Exchanging Spatial Systems Technology: D. R. F. Taylor, ed., *Current Issues in Geographic Data Processing*, Carleton University, pp. 70-91

Wynar, B. S. 1980, *Introduction to Cataloging and Classification*, 6th edition, Librairies Unlimited, Littleton, Colorado, pp. 400-402

Yan, Joel Z., Witiuk, Sid and Fisher, Terry. 1977, Exchanging Spatial Systems Technology: Related Issues and a Case Study, D.H. Douglas, ed., *Applications of Geographic Information Processing*, University of Ottawa, pp. 41-50

Zarzycki, J. M., Harris L. J. and Linders, J. G. 1975, Topographic-Cartographic Data Base and Automated Cartography in Canada: *Proceedings of the Commonwealth Survey Officers Conference*, Ministry of Overseas Development, Paper J 2, pp. 1-9

THE ARCHIVING OF COMPUTER CARTOGRAPHY

Louis Cardinal and Betty Kidd
National Map Collection
Public Archives of Canada
Ottawa, Canada

ABSTRACT

The cartographic archivist, throughout time, has had to adjust to technological changes in map production - for example, the effects of the printing press, lithography, and photography on mapping. Undoubtedly, the greatest challenge of the latter half of the twentieth century is the archivist's response to complex computer-based cartographic information systems. This response necessitates the evaluation of traditional methods for the appraisal, acquisition, control, and provision of public service to cartographic records, and the development of methods as required. Most of all, the archivist must quickly acquire an in-depth understanding of system design, the purposes of a system, and the interrelationship between these. This will be complicated by the vast number of unrelated systems, the nature of data bases and computer files which can be so easily erased and updated, and the lack of realization to date that such information may be of significant historical importance.

IGDMS
An Integrated Geographic Data Management System

Timothy L. Nyerges
C. Stephen Smyth
Smyth Associates, Inc.
200 Central Credit Union Plaza
PO Box 3406
Bellevue, WA 98009

ABSTRACT

Many references in the geographic information systems literature report special requirements for geographic data management that are not fulfilled by business data management system software packages. Although business data management systems can be used to manage geographic data, the majority are not well suited to the task due mostly to single data model conceptiona as well as a lack of storage structures for efficiently handling continuous geographic space. The EWorld™ Integrated Geographic Data Management System created by Smyth Associates, Inc. has been developed specifically to support geographic data management. The system is composed of four sub-system data managers: quad tree, simplex, relational, and unstructured. Rather than forcing an analyst to view the solution to a problem in terms of a single data model, a choice of data model(s) can be made most suited to the data. The sub-systems are described and application examples are cited.

APPENDICES

Steering Committee

Technical Program Committee, Editorial Committee

Acknowledgements

Author Index

STEERING COMMITTEE

D.R. Fraser Taylor
Conference Chairman
Carleton University

DIRECTORS

Tom Wilkinson
Conference Director
Carleton University

Gordon Deecker
Vice-Chairman and
Director Accommodation and Registration
Statistics Canada

Barry S. Wellar
Director Technical Programme, Proceedings Editor
University of Ottawa

David Rothwell
Director Exhibits
Compusave Inc.

Michael E.H. Young
Director Tours and Social Programmes
Energy, Mines and Resources Canada

David G. McKellar
Director Printing and Publicity
National Defence Canada

TECHNICAL PROGRAM COMMITTEE

Barry S. Wellar
Program Director
Department of Geography
University of Ottawa

Program Coordinators

Oavid Douglas
(Special Sessions)
Department of Geography
University of Ottawa

Henry Castner
(Education Track)
Department of Geography
Queen's University

Tom Poicker
(Problem Analysis Track)
Department of Geography
Simon Fraser University

Connie MacDonald and Ian Crain
(Applications Track)
Lands Directorate
Environment Canada

Ray Boyle
(R&D Track)
Electrical Engineering
University of Saskatchewan

Dick Groot
(Integrated Systems Track)
Surveys and Mapping Branch
Energy, Mines & Resources Canada

Program Advisors

Hartmut Ziemann
National Research Council
of Canada

John Kerr
Natural Resources, Ontario

Robert Aangeenbrug
University of Kansas

Lary Brown
Canada Centre for Remote Sensing

Robert Le Blond
Industrial Development
Research Centre, Canada

Nick Chrisman
University of Wisconsin - Madison

Joe Arbour
Environment Canada

Terry Fisher
Environment Canada

SYMPOSIUM PROCEEDINGS

Barry S. Wellar, Editor

David Douglas, Advisor to Editor

ACKNOWLEDGEMENTS

In addition to support provided by individuals, Auto-Carto Six was the beneficiary of financial, administrative and in-kind professional support provided by a number of agencies and individuals.

The following institutions warrant being acknowledged for the extraordinary assistance which they gave in support of the Program and these Proceedings: CARLETON UNIVERSITY and THE UNIVERSITY OF OTTAWA; Department of Energy, Mines and Resources Canada; Statistics Canada; Environment Canada; Canada Mortgage and Housing Corporation; Department of Fisheries and Oceans; National Capital Commission; Department of Regional Industrial Expansion; Department of Indian and Northern Affairs Canada.

The Steering Committee, on behalf of Symposium attendees and readers of this volume, also wish to acknowledge the contribution of the following institutions and agencies.

Agriculture Canada
Altek Corporation
American Congress on Surveying and Mapping
American Society of Photogrammetry
Associated Engineering Services Ltd.
Association of Canadian Map Libraries
Bausch and Lomb
Canada's Capital Vistors and Convention Bureau
Canadian Association of Geographers
Canadian Cartographic Association
Canadian Forestry Equipment
Canadian Institute of Surveying
Carriere Reynolds
Carto-Quebec
Collins & Moon Ltd.
Commercial and Industrial Development Corporation of Ottawa/Carleton
Commission de toponymie, Quebec
Compusave Inc.
Computer Graphics World
Department of Communications Canada
Department of National Defence Canada
Digital Equipment of Canada Ltd.
Dipix Ltd.
Eikonix Corporation
Exhibit Express
Gentian Electronics Ltd.
Geo-Spatial Research Corp.
GIMMS
Graphic Controls Canada Ltd.
Gregory Geoscience Ltd.
HASP Inc.
Health and Welfare Canada
Hill Security
Industry, Trade and Commerce
Intergraph Systems Ltd.
International Cartographic Association
International Development Research Centre Canada
Kern Instruments of Canada Limited
Kingsberg-Syscan
Laser-Scan Laboratories Limited

National Research Council of Canada
Northway-Gestalt Corporation
Omnitech Inc.
Ontario Centre for Remote Sensing
Ontario Institute of Chartered Cartographers
Ontario Ministry of Natural Resources
Optronics International Inc.
Photosur Inc.
Prime Computer of Canada Ltd.
Public Archives Canada
Queen's University
Simon Fraser University
Sir Sanford Fleming College
Societe de cartographie du Quebec
Spatially-Oriented Referencing Systems Association
Statistics Canada
Systemhouse Ltd.
Technicom Inc.
University of Alberta
University of Ottawa
Urban and Regional Information Systems Association
U.S. Bureau of the Census
U.S. Geological Survey
Wild Leitz Canada Ltd.

Ahn, J. 444
Alvo, M. 493
Armstrong, Marc 309
Bélanger, J.R. 524
Benson, J.A. 179
Bernard, A. 90
Bonnette, Barbara 329
Borgstede, Roy 410
Boss, Daniel 411
Bouillé, François 202
Bradley, Ross 117, 135, 570
Brassel, Kurt E. 416
Broekhuysen, Martin 15
Broome, Frederick 184
Bruger, Dominique 147
Cals, Eric 560
Cardinal, Louis 627
Cebrián de Miguel, J.A. 475
Chamard, R.R. "Sky" 541
Chow, W. 355
Colville, David 588
Cromley, Robert 435
Darcos, J. Claude 560
Dell'Orco, Pietro 299
Donnay, Jean-Paul 25
Douglas, David 2, 616
Dutton, Geoffrey 15, 186
Edelen, Robert 109
Falcidieno, B. 426
Forbes, D.K. 452
Franklin, Randolph 230
Freeman, H. 444
Gambaro, C. 426
Genest, Denis 484
Ghiron, Marco 299
Gibson, J.R. 165, 536
Gimblett, Randy 33
Gold, Christopher 412

Goldberg, M. 493
Goodenough, D.G. 598
Green, F.C. 85
Grelot, Jean-Phillippe 539
Guptill, Stephen 563
Hescock, Jonathan 178
Hobbie, Dierk 408
Hodler, Thomas 413
Holroyd, Michael 241
Honsaker, John 277
Hopkins, Lewis 309
Howarth, Philip 504
Itami, Robert 33
Jeansoulin, Robert 560
John, Rob 75
Karam, G. 491
Karonen, Olli 374
Kashyap, R.L. 243
Kerola, Penti 374
Kesik, A.B. 514
Kidd, Betty 627
Kinnear, Christine 177
Kiriakakis, Zissis 416
Kissling, C.C. 454
Korporal, Kenneth 605
Lam, Nina 319
Li, Shie-Yeu 44
Maher, Robert 588
Mark, David 288
McColl, W.D. 165, 536
McMaster, Robert 267
Meixler, David 177
Mitchell, K. 454
Mitjonen, Matti 374
Moellering, Harold 53, 319
Moore, Roger 331
Moore, W.C. 335
Morrison, John 44
Mruk, Stephen 384
Muller, Jean-Claude 277

Navarro, A.A. 549
Neville, R.A. 165, 536
Nyerges, Timothy 628
Ommaney, Simon 504
Ommer, Rosemary 491
Oomen, John 243
O'Neil, R.A. 165
Orivuori, Esko 374
Ozone, Mohammad Id 561
Palimaka, J.J. 598
Pellerin, Gervais 99
Piquet-Pellorce, D. 90
Plunkett, G.W. 598
Puittinen, Dave 514
Read, Chung Hye 127
Reed, Carl 409, 559
Rigby, Douglas 588
Rodrigue, M. 580
Rugg, Robert 211
Ruys, Joseph 155
Schiro, Richard 252
Schneier, Jan 155
Shelberg, Mark 319
Shirtliffe, Glen 65
Siekirska, Ewa 464
Sinigaglia, P. 426
Smyrnew, John 443
Stover, Richard 398
Strewig, Larry 155
Tamminen, Marku 374
Tarvydas, Albin 240
Taylor, M.J. 454
Teng, Apollo 348
Thompson, L. 580
Thrift, N.J. 454
Till, S.M. 165, 536
Trainor, Timothy 184
Vogel, Stephen 184
Wild, S. 454
Williams, Glen 252

Wilson, Paul 221
Wiltshire, Steven 365
Wood, Bruce 616
Wood, Cliff 491
Yan, Joel 135
Yoeli, Pinhas 262
Zastrow, Joseph 412