LAND INFORMATION MANAGEMENT
GEOGRAPHIC INFORMATION SYSTEMS
ADVANCED REMOTE SENSING

Edited by

Ewan G Masters & John R Pollard

Vol .1 PROCEEDINGS

LIM & GIS CONFERENCE

Sydney 1993

proceedings sponsored by

ESRI AUSTRALIA GEOGRAPHIC INFORMATION SYSTEMS

MONOGRAPH 1

SCHOOL OF GEOGRAPHY

MONOGRAPH 15

SCHOOL OF SURVEYING

THE UNIVERSITY OF NEW SOUTH WALES KENSINGTON NSW AUSTRALIA 2033

Proceedings of Conferences

on

LAND INFORMATION MANAGEMENT
&
GEOGRAPHICAL INFORMATION SYSTEMS

and

ADVANCED REMOTE SENSING
Coming of Age - 21 Years after Landsat

SYDNEY - AUSTRALIA

19-23 July 1993

Edited by

EWAN G. MASTERS
&
JOHN R. POLLARD

School of Surveying
School of Geography
Centre for Remote Sensing and GIS
The University of New South Wales
PO Box 1 Kensington 2033
AUSTRALIA

The School of Surveying
The University of New South Wales
PO Box 1 Kensington NSW 2033
AUSTRALIA

Published 1993

National Library of Australia
Card Number and I.S.B.N.
0 85839 064 7

ESRI AUSTRALIA

GEOGRAPHIC INFORMATION SYSTEMS

THESE CONFERENCE PROCEEDINGS WERE PRODUCED WITH THE SPONSORSHIP OF

ESRI AUSTRALIA

Hosted by

The Schools of Surveying and Geography
and
The Centre for Remote Sensing and GIS

The University of New South Wales

DISCLAIMER

SPONSORS

The organizing committee expresses their appreciation to the undermentioned sponsors for their support of the conferences

ANSETT Australia
AURISA New South Wales Chapter
AUSTRALIAN CENTRE FOR REMOTE SENSING
AUSTRALIAN SPACE OFFICE
CSIRO - Space Science
ESRI Australia
GENASYS II
GEOVISION
LAND INFORMATION CENTRE (NSW)
REMOTE SENSING PHOTOGRAMMETRY Australasia
SPOT IMAGING SERVICES
THE UNIVERSITY OF NEW SOUTH WALES

Organising Committee

B.C. FORSTER
E. G. MASTERS
J. R. POLLARD
A.J. ROBINSON
A.S. SKIDMORE
J. C. TRINDER
M. WINDLE

PREFACE

These conferences in Land Information Management and Geographical Information Systems and Advanced Remote Sensing come 21 years after the launch of Landsat. We believe this is a significant occasion and hope the format of parallel conferences will provide a major event to mark the occasion. The Schools of Surveying and Geography and the Centre for Remote Sensing and GIS have many years of involvement in Land Information Systems, Geographic Information Systems and Remote Sensing. The Conference on Land Information Management and GIS follows the very successful Conference on Land Information Management held at UNSW in July 1991 and the occasion of 21 years of successful Landsat operations warrants a special conference. The running of parallel conferences, we believe, highlights the growing synergy between disciplines which in the past have been separate and which will become inseparable in the future.

We have been fortunate in persuading the distinguished overseas experts Professors D. Rhind and P. Burrough to present keynote addresses for the conference on Land Information Management and GIS and Professors J. Richards, J. Curlander, J. Huntington, H. Houghton and P. Curan to do likewise for the Advanced Remote Sensing Conference. It is our hope that this collection of papers will provide a useful future reference of the status of LIM, GIS and Remote Sensing in 1993.

E.G.M.
J.R.P.
June 1993

CONTENTS

GIS APPLICATIONS IN THE TRANSPORTATION INDUSTRY

Michael Avery
Senior Associate
Booz•Allen & Hamilton
Level 7, 7 Macquarie Place
Sydney, NSW, 2000 Australia

Brian Stearman
Principal
Booz• Allen & Hamilton
4330 East West Highway
Bethesda, Maryland, 20814 USA

Bruce Brong
Brong & Associates
Boonesboro, Maryland USA

ABSTRACT

This paper summarises recent GIS experience in the transportation industry. The commonly used term GIS-T will be used in this paper to refer to GIS applications in the transportation industry. The paper consists of 4 sections:

Section 1	–	Who Are GIS-T Users?
Section 2	–	Significant Issues Related to GIS-T
Section 3	–	Sample of Current Applications of GIS-T
Section 4	–	Future/Potential Applications of GIS-T

This paper is not an inclusive world-wide review of all GIS transportation applications. It does provide representative uses of GIS in the transportation industry. A working knowledge and background of GIS is assumed.

SECTION 1 - WHO ARE GIS-T USERS ?

A User Profile

There are literally thousands of GIS-T applications currently operating around the world. Over 80% of them are government or quasi-government agencies. This group includes:

- direct government agencies like federal, state or local governments responsible for transportation and public works; and
- quasi-government activities including public utilities and educational institutions.

The investment to date of GIS-T in the private section is dwarfed in comparison to that of the public sector. Private sector applications are fewer and are concentrated in these industries:

- oil production and distribution;
- location and routing of vehicles, trucks, aircraft and ships;
- transport of freight and passengers; and
- forecasting of demand for services and goods.

Virtually all federal transportation agencies in industrialised countries have a development or research project dedicated to GIS-T. In general, GIS-T efforts are not coordinated. It is not uncommon for one federal agency to be researching GIS-T to better co-ordinate public transportation needs, while a second department is using a different GIS-T approach and technology to research environmental and transportation related problems. Fortunately, there are indications that this situation is changing.

An important report was produced in August 1991 for the National Co-operative Highway Research Program. The report is entitled "Adaptation of Geographic Information Systems for Transportation" by Alan Vonderole, Larry Travis, and Robert Smith, of the University of Wisconsin. Some key results of the interim report that are relevant for this paper are:

- In most states the key impetus for GIS-T was from interested staff, rather than top management. In the majority of cases GIS-T projects began in the planning department. In some cases, GIS-T began in engineering, management information systems or geodetic departments. There are examples of GIS-T being developed simultaneously in multiple departments. Two of fifty states have little interest in GIS-T.

- Formal strategic plans for GIS-T are rare and usually non-existent. Lately, it is becoming more common for states to participate in statewide GIS-T development efforts, that involve non-transportation agencies.

- The most common applications of GIS-T are inventory, accident analysis and pavement management.

- At least twenty different combinations of hardware and software are being used or tested.

- There are three major impediments to GIS-T development projects: manpower, funding and top management support.

SECTION 2 - SIGNIFICANT ISSUES RELATED TO GIS-T

The design, development, and implementation of a GIS can require a significant resource commitment and should be carefully planned. Based on the experiences of other transportation agencies, various issues have been identified as common to any agency who undertakes the implementation of a GIS.

This paper describes representative GIS-T projects in a later section. To better describe the significant issues of GIS-T design and development, this section begins with a description of a large, comprehensive design and implementation project.

A Multi-Participant GIS Project

In the Indianapolis, Indiana (USA) area one of the largest multi-participant GIS projects yet undertaken is moving toward completion. The goal of the project is the construction of a common, comprehensive land-related database. The project will link the mapping information of over two dozen users from city and county agencies, public utilities, private industries and universities.

The Indianapolis Mapping and Geographic Infrastructure System (IMAGIS) will provide a geographic-based information system for planning, engineering, utility and related public service activities.

This computerised system will improve use of city and county information in a number of ways. The system will transform 8.3 million map documents used by over 2,600 employees in the 28 participating organisations into a common database of shared information. Previously, each of the organisations guarded their own set of maps and records. These documents had been collected and updated over many decades and were prone to error, duplicated to some extent in different organisations, difficult to read, and sometimes inaccessible. Because of the lag time in updating the material, many of the documents were out of date. A review showed that US$9 million was spent per year on existing manual land-related tasks such as creating, maintaining and verifying records. The projected cost of the four year project is estimated at US$9.2 million.

The project has four phases. The first phase, "Project Design and Preparation" consists of:

- analysing each participant's needs;
- defining the system's functional requirements;
- developing a strategic implementation plan;
- conducting a cost/benefit analysis;
- preparing a funding and cost allocation plan; and
- purchasing and starting up the hardware/software configuration.

The second phase, "Aerial Photography", provided primary land-based data. The next phase, "Digital Base Map Creation", provided a means of inputting each participant's specific information. Among the different participants, over 50 database elements were defined in categories such as: boundaries/areas, control, labels/attributes, land use, natural features, facilities, rights-of-way, structures and transportation. Map accuracy was a significant issue since it has a greater impact on base map cost than any other consideration. The project team took the view that there will never again be an opportunity to obtain accuracy at a lower cost; therefore, they specified a +/- 2 feet accuracy. The final phase, "Digital Facilities/Infrastructure Creation", allows individual participants to use the database as a source for different application programs.

The IMAGIS consortium purchased only the central mapping hardware/software, the original land base creation and certain land base updates. Each participating organisation was responsible for the acquisition of its own connecting hardware and data manipulation software. The consortium's hardware, furnished by Digital Equipment Corporation, is housed at the Indiana University Purdue University at Indianapolis (IUPUI).

The conversion process of data to digital form proceeded along several fronts simultaneously: stereo compilation, data gathering, output and quality control. A representative six square-mile area was first converted. The total coverage area was broken into discreet areas. Only after one area was complete was the conversion of another area begun.

Data integrity was a critical issue. To address it, the implementation plan outlined procedures for construction and maintenance of the land base. All participants can access and display land base information. Read-only access, however, is granted to those requiring access to the data but not directly responsible for maintenance of the data. Each participant is responsible for the accuracy and integrity of the information in its area of expertise. The data types unique to each participant's applications are layered onto the land base. Each group can allow access to its information by others.

Among the organisations using the land base to improve their operations are the Indianapolis Department of Public Works and the Indianapolis Department of Transportation. The Department of Public Works will access the IMAGIS database as it tracks complaints, generates and tracks work orders, and programs and records routine sewer maintenance and inspection activities. The Department of Transportation is adding the type and cost of maintenance work performed to the database. DOT's maintenance management system produces reports tying work quantities and costs to specific links defined in the IMAGIS land base. The record of work performed is available for others to consult.

Since this project used the resources of and provided benefits to over two dozen participants, the problem of who pays how much demanded careful consideration. The consultant hired by the consortium devised a cost allocation plan that allows all users, regardless of size, to participate in the project. The cost allocation plan specifies that government participants assume 80 percent of the automated mapping cost and utility groups assume 20 percent. The plan allows any identified participants to connect to the IMAGIS infrastructure network when they become financial participants. Some users connected as far back as 1987 and some still have not connected. The cost of participation is also allocated according to the user's mapping accuracy requirements. Since map accuracy requirements have a direct bearing on the cost of creating the land base, those groups with higher accuracy needs contribute more toward the cost of creating the land base.

The IMAGIS shared participation concept provides an important model for organisations contemplating the benefits of a GIS-T but awed by the magnitude of both the work involved and the expense of implementation.

Top Management Support is Required

There are three major problems facing GIS-T design and development efforts today. Lack of resources, funding and top management support. They are really one and the same. If top management is committed to a project, funding and resources become available.

A very basic GIS-T can be implemented for less than AUS$25,000, but the cost can easily escalate to millions of dollars. With that price tag, it is important that agencies carefully evaluate their requirements. A decision to implement GIS-T should not be made on the basis of "it seems like an interesting project". Each agency must define what its expectations are of GIS-T and clearly decide if GIS-T is cost-effective.

GIS-T has a wide range of capabilities and offers potential to be useful in a variety of fields and situations. This flexibility allows agencies to better justify GIS-T purchase, development and implementation.

Consider the following example. The Pavement Management System (PMS) group may decide that graphical display and analysis of data is a natural and useful enhancement to their basic PMS software. The graphical display of data in and of itself is not sufficient justification for a GIS-T.

There are, however, additional groups in the same agency that could also benefit from GIS-T. The GIS-T Project Manager could involve representatives from those groups to share the development and implementation costs of GIS-T. In the case of this agency, the Land Records, Maintenance Management System, and Accident Analysis group may be interested in the graphical and analytical capabilities GIS-T offers.

There is a valuable lesson to be learned from this example. Although a single interest group may not justify the costs of GIS-T, within a typical agency, there are likely to be others who could share the costs. If properly planned and co-ordinated, the needs of all participatory groups can be met within the same scope of work. This may slow the development effort to satisfy all user requirements, But if top management support is the key, this is the preferred method.

The Need for Spatial Referencing is a Key Decision

GIS-T graphically displays data on maps and different maps may be layed on top of one another. In many applications, the display of data is location sensitive. For example, analysis of the relative locations of roads to rail lines may be important. A common positioning system is required to ensure roads are displayed in the correct relative position to the rail lines on the same map. When GIS-T attempts to match road condition and location, accident, sign inventory, rail lines, and any contours from base maps, on a single integrated map, a common referencing system must be used.

Many of the maps and data locations are already contained in the existing records of an agency. At the time when the records were developed, each user group in the agency used a referencing system appropriate to them. Some examples of common referencing systems include longitude/latitude (long/lat), AGM (in Australia), kilometre post markings and names of intersections. Unfortunately, research in the USA indicates it is not unusual for a state department of transportation to have up to ten different referencing systems within their organisation.

The decision that needs to be made is how these different spatial referencing systems will be co-ordinated. There are three very general approaches:

- Don't bother to reference the data to specific map locations. Where this is possible, the co-ordination problem disappears, since locations are not considered critical to the application.

- Use a common referencing system to co-ordinate each of the maps. If the base maps for accidents, road locations and rail lines all reference longitude/ latitude, the common reference system would be in place and no conversion would be necessary to overlay all maps. This alternative requires conversion of existing records to a new co-ordinate system.

- In most cases however, various referencing systems have been used, and the effort to revise existing records is prohibitive. In those cases, a conversion process of existing records will be required. For example, the GIS-T conversion programs would translate the AGM co-ordinates used for the road locations, to the long/lat co-ordinates used for rail line maps.

Keep User Expectations Realistic

GIS-T has become a "buzzword" in many transportation agencies around the world. Unfortunately, this has raised expectations of many, who do not fully realise the capabilities and limitations of GIS-T.

GIS-T, as with any other information system, is not able to make decisions. GIS-T will perform some analysis, display data and provide information that was not previously available. It will not eliminate the need for managers to decide and act.

Expectations must be differentiated from requirements. Both expectations and requirements must be realistically discussed before a project begins. In some cases, it will be possible to satisfy all expectations, but at a very high cost.

It is the responsibility of the GIS-T Project Manager to evaluate users' expectations and estimate the effort and costs required to satisfy them. As a Project Manager, it is better to not falsely raise expectations, than to have your project approved based on promises that cannot be fulfilled.

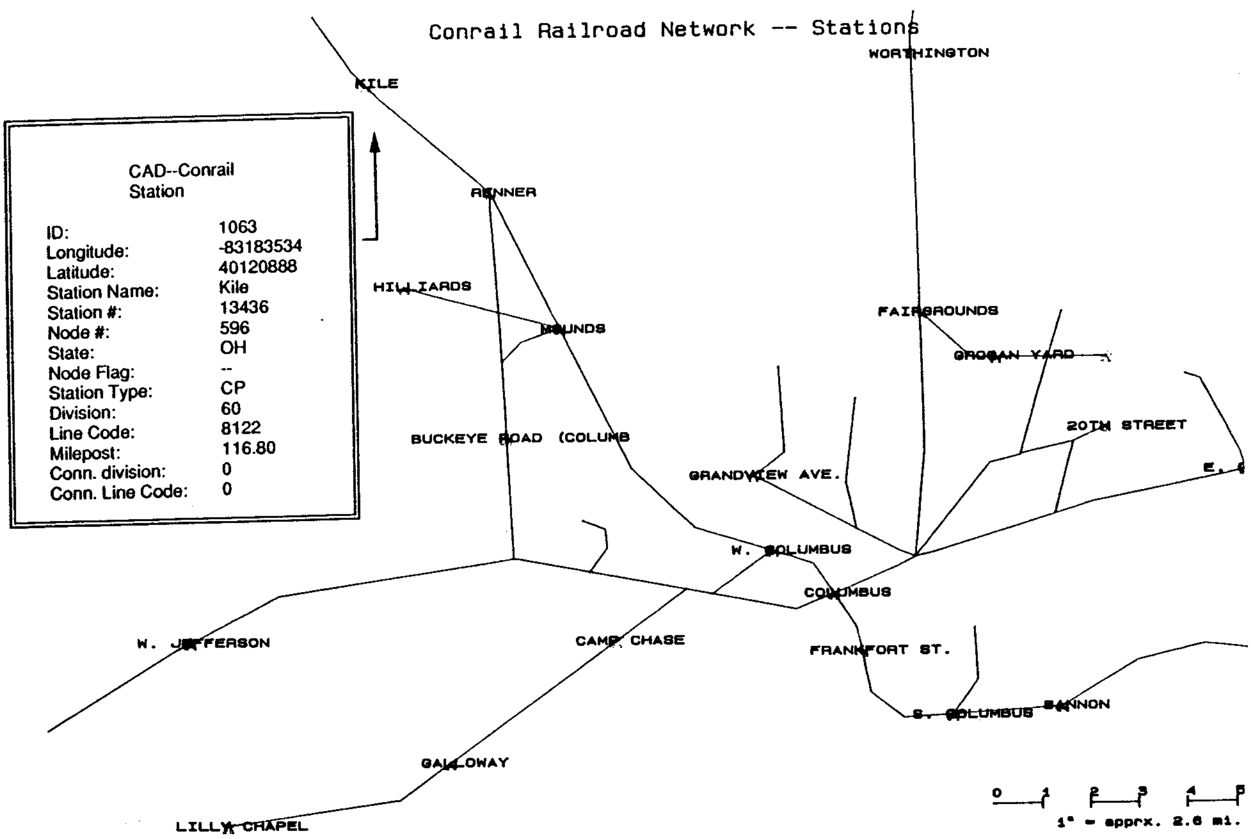
SECTION 3 - A SAMPLE OF CURRENT GIS-T APPLICATIONS

ConRail (USA)

ConRail is one of the largest railroads in the USA. In recent years it has become a major carrier of hazardous materials. Some of its major routes are through densely populated areas of the north-eastern USA.

It is currently applying GIS-T technology to analyse routing for hazardous materials. The result of its analysis is selection of the routes that minimise safety risks while minimising costs. Early in the project, the decision about co-ordinate systems was considered. The decision at that time was to not spatially reference rail lines and population centres to a long/lat co-ordinate system. It was not necessary to perform the routing analysis.

A sample graphical display is presented below. ConRail is using Caliper Corporation's TransCAD software, on an IBM-compatible computer with a 300 MB HDD.



Roads and Traffic Authority, NSW, Australia

The Roads and Traffic Authority (RTA) of New South Wales is the state roads agency. They have nearly 9000 employees and an annual budget of AU$1.6 billion. The RTA has recently undertaken the design and development of a GIS-T. The main purposes of the system are display and analysis of data for network performance, condition and use as well as integration of a total asset management system.

The system currently uses AGM co-ordinates, but development efforts in the near future may be converted to long/lat. The RTA uses Genamap software with Oracle data bases, running on a HP Series 700, Unix based computer.

Wisconsin, USA, Department of Transportation

The Wisconsin Department of Transportation (WisDOT) has one of the more advanced Geographic Information Systems of any of the DOT's in the USA. They have completed the digitisation of the state's highway system on a 1:100,000 scale and are operating a Pavement Management System (PMS) based on their pilot GIS-T.

WisDOT is building a demographic information layer into the GIS-T to help the state legislature with the task of redistricting the voting districts. WisDOT has a staff of 24 people who develop and present training courses and publish a newsletter devoted to issues affecting GIS-T. Currently, WisDOT uses ARC/Info software running under the UNIX operating system on 25 networked workstations.

WisDOT began the first phase of its GIS-T development in 1986 when it started a series of exploratory projects. The aims of the Technology Acquisition phase were to introduce GIS-T concepts and products to the WisDOT community and to gain an in-depth understanding of the resources necessary to implement GIS-T across the Department. The projects in this first phase included an evaluation of the state of practice in GIS-T; an evaluation of GIS-T's applicability to the needs of the Department; a feasibility project designed to gain experience with the ARC/Info software; a pilot project which explored the use of GIS-T as an application development platform and the acquisition of the initial hardware, software and databases.

The second phase of the WisDOT's GIS-T development began in 1990 and is defined as Infrastructure Development. This phase consists of putting together the people, procedures, training and standards upon which subsequent application development can be based. The Department has written a development plan which states the GIS-T objectives and outlines the tasks and resources necessary to evolve GIS-T prototype applications into a major service initiative.

The objective of WisDOT's GIS-T, is "to provide a single, coherent technology which can be used to integrate, analyse or report on all spatial data collected and maintained by the Department". To accomplish this objective, WisDOT plans to establish GIS-T capabilities to support its varied requirements, promote sharing of spatial data, minimise redundant data collection and provide a single integrated spatial information environment.

North Carolina, USA, Department of Transportation

The North Carolina Department of Transportation (DOT) has been developing their GIS-T for two years. They began the process three years ago with a software evaluation and settled on ARC/Info software. Currently, they are digitising the state's transportation layer at a 1:24,000 scale. They have been working on the digitised maps for two years in conjunction with the US Geological Survey. To produce the digital files of the maps, the DOT is using an Intergraph MGE workstation running the USGS's RETSAM software. They expect the total digitising (by the two agencies) to last three years.

The DOT plans a variety of applications for its GIS-T. To display their information and support the GIS-T applications, the DOT is running ARC/Info software on Sun Microstations. They have already created a demonstration application of optimum corridor selection. By collecting and displaying information about land usage, trip origin and destination, population centres, zoning restrictions, topography and more, the application can make preliminary selections about proposed highway corridors. The users can then modify the application's selections and produce an optimum corridor. To support this application, the DOT worked with another state agency which is digitising the necessary environmental information.

The DOT is also working with the Federal Highway Administration on an application to preserve potential future highway corridors. In this application, the DOT will use the GIS-T to identify potential corridors. Then the DOT will work with other agencies to limit development in these corridors. By limiting certain kinds of development, the DOT will hold down the future cost of right-of-way acquisition.

The DOT's original plan for its GIS-T identified three uses to justify the implementation of a system; pavement management, accident location analysis and general information queries. The DOT however, has only prepared models of the pavement management system it intends to create and has begun data collection for it. The development of the accident location analysis application is in progress. Full implementation, however, awaits the completion of the map digitisation, although the state police have accident statistics in a separate database. The DOT has created a few demonstration applications and planned the general query system, but has not developed the actual application yet. This general query system is intended to provide users with information about highway geometrics, maintenance histories and traffic count data, among other items. As each county's highway system is digitised, the DOT will develop different applications in parallel.

DOT officials have learned many lessons from their efforts to establish a GIS-T. They have been fortunate to have had strong support from their top management and consider that support a crucial factor in their success so far. The state's governor recently signed legislation establishing a statewide GIS co-ordinating committee. This committee will promote the growth of GIS-T in different state agencies by establishing standards, enhancing co-ordination between agencies and developing policies for the maintenance of data. Each state agency will retain ownership and perform maintenance on its data, but data will be available to all state users.

Obtaining and retaining an appropriate staff has been a significant problem. The state's personnel agency has been reluctant to create a special classification for the DOT's GIS-T trained people and the DOT has made use of university students and faculty to accomplish some development work. Turnover coupled with the significant learning curve of the software have combined to slow development. One DOT official suggested that agencies contemplating GIS-T development should talk to other agencies which have gone through the process. Members of the agency should make visits to counterparts in experienced agencies.

Before deciding on software, the agency should insist on extensive testing of different vendors' products. This means not just demonstrations of software by vendor representatives but obtaining working copies which are put through a series of tests. It is critical to obtain top management backing for the effort. It is equally important to retain this backing by producing periodic demonstrations to maintain management's interest, conviction and enthusiasm.

Ontario, Canada, School Board

The London (Ontario, Canada) Board of Education is using a GIS to help manage the transportation of its students to school. The Board is developing an Integrated Transportation Optimisation and Planning System (ITOPS). The Board is responsible for transporting over 12,000 students to school each day. The students use either the Board's fleet of vehicles or the city's municipal transit system. Co-ordinating the movements of the students to one of more than 70 schools in the system presents the Board with serious logistical and budgetary problems.

The GIS-T will help the Board deal with these problems by providing digital data storage and analysis functions. ITOPS will help transportation planners to:

- model scenarios;
- calculate the most efficient school districts and bus routes;
- allocate vehicles; and
- anticipate changes in the demographic and geographic patterns of the school population.

The information provided by the system will allow the Board to judge policy, plan for transportation and capital costs, prepare alternative policies for discussion and respond to public inquiries.

The system consists of three modules. The Corporate Data Maintenance Module will maintain the database. The street network files will be kept up to date using input from subdivision plans furnished by the city. The database will contain street names, address ranges and street graphics. Also included will be tax assessment data, student records and data about schools and study programs. The Transportation Module will help plan the bussing program. In this module, officials can determine a student's eligibility for transportation, assign students to bus stops or public transit and maintain bus stop coverage and vehicle files.

A sub-module will optimise bus routes. The sub-module will determine the most direct route for each bus, taking into account the number and capacity of the buses, the location of stops and the number of students at each stop. The sub-module will allow planners to test different scenarios. The system will automatically generate route maps for the bus drivers, notification cards for the students and data summaries for the Board.

The Planning Module will help school planners anticipate changes in the school population. Information in the database will enable planners to project demographic trends, create multiple school districting scenarios and simulate the effect of policy changes. The module will provide for automatic re-districting when new sub-divisions are built or the demographic composition of the population changes.

The ITOPS runs on the Board's VAXstation 3100E. It uses ARC/INFO GIS software and Oracle as the relational database. Initially, three terminals will be able to access the VAXstation but the Oracle database will be accessible to any of the more than 100 terminals.

IVHS in Europe

GIS-T in Europe starts with an information advantage that others around the world can envy. Most states in Europe have long-established mapping agencies, some going back more than two-hundred years. These mapping agencies are responsible for the maintenance of national databases at the 1:50,000 scale. Federal mandates, in some countries, require that this database be updated every two years. One requirement which aids this update process is the construction obligation to register a complete set of detailed plans with the mapping agency before an occupancy or use permit will be issued. The availability of these databases make European-wide projects like the following one easier to implement.

A joint project, sponsored by the Council of Europe and including the Damlier-Benz vehicle manufacturer, is developing the hardware and software necessary for intelligent vehicle/highway systems (IVHS). Such systems combine sensors embedded in the roadway, computerised cameras, global positioning systems, two-way radio data communications, on-board vehicle information processing and display systems, and a central traffic information system to guide vehicles through urban traffic. The central database for the system is held in a GIS-T which contains highway and street geometrics, physical and legal restrictions to movement, routing information, optimisation algorithms and vehicle characteristics.

The roadway sensors and computerised cameras report to the central system on current traffic conditions. The on-board systems use global positioning to determine an individual vehicle's position in relation to the highway system. The central system monitors traffic conditions, determines optimum routings and issues advisories which are displayed directly on-board the vehicles moving through the highway system. The central system can also plot the most expeditious route for a particular vehicle upon request and communicate it to the driver. IVHS, when fully operational is expected to result in smoother, faster traffic flows, improved transportation safety and reduced pollution.

This Council of Europe project, the Multi-purpose European Ground Related Information Network (MEGRIN), is in its pilot phase. Preliminary testing of the concepts and prototypes of the hardware have been conducted in 12 cities in seven European countries. The necessary GIS-T database and application programs for positioning and display are nearing completion. The central processing hardware and the software to analyse and optimise traffic information is under development. The on-board vehicle processing and display units are expected to be available in 1994.

The need for support from country governments is a critical factor for the success of the project. The final unification of the European Economic Community holds out the hope that the necessary government support will be provided. The accurate geographic information collected by the mapping agencies of each country and made accessible by a GIS-T will help speed the development of IVHS in Europe.

China Railroads

Members of a World Bank evaluation team recently used GIS-T to help manage a pilot project for the China Ministry of Railways. This project was part of the Railway Investment Study and Coal Transport Study in preparation for an expected Bank loan. The aim of the pilot project was to develop railway models for coal transport and more general ones, and to lay the ground work for performing cost/benefit analysis.

In this project, TransCAD software operating on a Compaq 386 PC platform provided the user interface, data query and display. The TransCAD software was used to build a low resolution map and database of the railway network. A co-ordinated group of programs was used to analyse various scenarios. The network database contained some 400 links with their operating characteristics. The operating characteristics were imported from an existing Chinese database. An abstract representation of the eastern coast rail network was digitised by hand. The representation consisted of about 300 nodes but geographic accuracy was not considered important. Rather, the relative location of the nodes was used for the data query interface.

This database was used to calculate capacities and running times under different conditions. Certain information from the database was exported to outside programs for processing. One outside linear programming program performed a flow forecast of railway traffic; another program performed economic analysis and net present value calculations on both railway improvements and cargo values. The results of these outside programs were imported into TransCAD for display and report preparation. The upcoming project is expected to include enhanced versions of the pilot GIS-T to help with the cost/benefit analysis.

Two significant drawbacks of the pilot were noted. Firstly, the PC platform proved to be severely limited in its capacity to run the optimisation program. Secondly, the GIS-T software was restricted in its data export and import capabilities. For the expected pilot, a workstation or minicomputer is considered a more appropriate platform.

SECTION 4 - FUTURE/POTENTIAL GIS-T APPLICATIONS

Although GIS is widely used in other fields, applications in the transportation industry are in their relative infancy. Future GIS-T applications will largely mirror those in other fields. These include:

- **Integration of GPS**

  Global Positioning Systems (GPS) locate vehicles, people, aircraft, trains and ships via satellite sensing. GPS sensors are in wide use around the world. As the satellite network develops and expands, the potential for more accurate, less expensive positioning expands.

  There are many efforts in the vehicle, train, aircraft and ship industries to integrate GPS and GIS-T technologies. The more sophisticated of these provide real time graphical display of movements of goods and people. The most common commercial applications are map displays in vehicles that use GPS or dead reckoning devices to track vehicle movements. These units can analyse routing to a predetermined point to minimise traffic congestion and travel time.

  In the future, GPS and GIS-T will merge into a single technology that will be transparent to the user.

- **Substantial Growth in the Industry**

  GIS-T use will increase by 100% in the next three years as more transportation officials recognise its capabilities. This growth will not be unique to GIS in transportation only. Additional hardware will be installed, and more sophisticated applications will be developed.

- **Hardware/Software Will Dictate Future Applications**

  GIS-T applications are limited by available hardware and software. As larger volumes of data can be stored and processed, transportation applications will more closely model "real life situations". It will be common in the future for transportation applications to analyse and display "real time" data in three dimensions, and for decisions to be made through artificial intelligence or neural networks.

  As hardware and software capabilities are enhanced, the need to supply complete, accurate and integral data will be more important. These needs will in part be met by Global Positioning Systems and related equipment.

- **GIS Will Become Transparent To the User**

  GIS-T is currently an automated tool to display and analyse information. GIS-T is not the answer that transportation officials seek. Rather, it is a necessary tool to provide answers to everyday problems.

  As GIS-T and related tools are developed, enhanced user interfaces will make GIS-T as a package, transparent to the user. Users will simply query the application and not need to be as familiar with how answers are produced. The net effect of this is that users will spend more time managing and less time learning to use the tools. GIS-T will be integrated into other information systems and gradually disappear as a separate and distinct application.

LAND INFORMATION SYSTEMS AT THE LAND INFORMATION CENTRE

By
George Baitch
and
Lloyd Pilgrim

1. INTRODUCTION

In the past two years a number of presentations relating to cadastral reform have been made at various fora. These presentations outlined progress relating to cadastral reform, from both a legislative and systems development aspect.

This paper intends to report the current status of cadastral reform and progress towards systems development. Further, an outline of future directions and options will be given.

2. REVIEW OF SYSTEMS DEVELOPMENT

2.1 The Concept

The systems development being referred to has been called the cadastral spatial referencing system (CSRS). This name was coined in the report which initially documented the requirements of a spatial referencing system (Toms & Baitch, 1989) providing a surveyed information management system. The terms of reference for this study were;

> "..*to update and complete the existing draft plan for the introduction into NSW of a spatial referencing system that will eventually yield 'absolute co-ordinates' for every cadastral corner in the state.*" (p. 1)

For the above report the following goal was developed:

> "*To develop an implementation plan for the establishment and maintenance of a Cadastral Spatial Reference System for New South Wales that will provide for positional accuracy adequate to needs of the community that utilises surveying services.*" (p. 12)

In order to achieve the above goal, five option plans were examined by Toms (*et al.*, 1989). These were:

- a. Progressive Co-ordination
- b. Historic Data Co-ordination
- c. Co-ordination by Resurvey
- d. Double Hybrid Model (Progressive - Historic)
- e. Triple Hybrid Model (Progressive - Historic - Resurvey)

The result of the Toms report was that the *Progressive-Historic Model* (plan d) was recommended as having the highest priority. This recommendation was accepted by executive management. This plan required the development of a cadastral co-ordinate layer in the survey control information management system (SCIMS) for the management of the coordinate information.

After the Toms report was presented the task of implementation began. It soon became evident that more data than co-ordinates of cadastral corners would be required to provide a genuine surveyed information management system. It was decided that rather than establishing the cadastral spatial referencing system as a "layer" of SCIMS, the two products were to be physically separated and thus allowing for;

both spatial and aspatial cadastral information to be managed in electronic form by a new independent management system known as the cadastral spatial referencing system.

2.2 The Momentum

In 1990 the State Government allocated $27.6 million dollars for the acceleration of a number of projects being undertaken by the Land Information Centre. One of these projects was in fact the introduction of a co-ordinated cadastre, otherwise referred to as the cadastral spatial referencing system (CSRS). This meant that external support was now in place for the development of the cadastral spatial referencing system, and so expertise from outside the organisation could be sourced.

Development of a separate data base model for the cadastral spatial referencing system began in late 1990 by designing the necessary data base tables which were to be sited on Land Information Centre computer facilities.

Further, a transfer mechanism was required to move the data from its source (i.e., cadastral surveyors) into the new data base tables. This activity was met by designing an electronic data transfer standard which became known as the neutral cadastral file (NCF) format. A consultant, was engaged to assist in analysing the elements and entities that were to be transferred into the survey data base by the neutral cadastral file format.

Once the transfer mechanism was developed, it was used for the capture of historical survey information in a series of projects that became known as the collaborative projects. These projects supplied a series of historic data sets that have been used for testing systems developments, as well as providing an opportunity for the private surveying sector to become familiar with creating and transferring cadastral data in digital form.

2.3 The Focus

By late 1991 a systems design consultant to the Land Information Centre had identified the necessity to develop a corporate survey information system to manage the flow of all spatial information of relevance to the survey system. This information system would incorporate four separate data bases;

- the cadastral spatial referencing system,
- the geodetic observation management system,
- the Board of Surveyors data base, and
- the Geographic Names Board data base.

It was proposed that these be combined into one effective overall management system. As a result, a primary logical design model was developed for the overall management system - the survey information system (SIS). The survey information system model allowed for all these data sets to have connectivity with the DCDB.

After the logical design model was developed (mid 1992) another consultant was engaged to develop a physical implementation model for the cadastral spatial referencing system. As part of this contract, a specification document for the development of the cadastral spatial referencing system was to be developed. At the same time the other components of the survey information system were being modelled and specified in conjunction with the cadastral spatial referencing system.

In order that the cadastral spatial referencing system met all the perceived user needs, and supplied all the necessary requirements for the survey information system, the amount of data to be managed became quite vast. The volume of data had increased from merely co-ordinates in 1989, to virtually all the information content of a deposited plan and more by 1992.

In late 1992 a user requirements survey and cost benefit analysis was commissioned for the cadastral spatial referencing system. Considering the many changes it had under-gone in the past three years, it was felt that the preliminary design should be tested against user needs. The terms of reference of the User Requirement Study was to:

- advise on the need to provide a mechanism to update the DCDB with survey derived information;
- investigate the need to provide spatial cadastral data during the development stage of land;
- identify the difficulties in matching survey derived information with the DCDB.

The predicted implementation costs were to be weighed against the community benefit in order to ensure that the project still provided a realistic function to the wider community, as well as the internal client (Land Information Centre).

The User Requirement Study indicated that there is not a clear understanding by many of the identified users of the benefits of a cadastral survey information management system though the benefits have been widely identified and reported (Aptech, 1986, Toms & Johnstone, 1989, and Angus-Leppan & Angus-Leppan 1990). The study indicated that the users who only wish to have an update mechanism for the DCDB can not see any justification for proceeding with the development of this system.

As of June 1993 the project status is as follows:

a. A pilot study of the proposed system is being carried out to trial data input and output availability and form.
b. Specification document is undergoing final in-house preparation.
c. User requirements survey and cost benefit analysis is in final draft stage.
d. Neutral cadastral file format is undergoing adaptation according to surveyor feedback.
e. Collaborative projects have delivered about 11,500 lots in electronic form (neutral cadastral file format), with a further 15,000 expected in the next 2 to 4 months.
e. Electronic data validation and quality control procedures are in place.

3. PRESENT LAND INFORMATION SYSTEMS PHILOSOPHY

3.1 Introduction

In considering the philosophy of land information systems at the Land Information Centre, it should be understood that there are existing and proposed systems. The digital cadastral data base (DCDB) is an example of an existing cadastral system, while the cadastral spatial referencing system is an example of a proposed cadastral system.

In order for the Land Information Centre to maintain its world leadership status in land information systems, we need to use existing products for today, but develop new products for tomorrow.

3.2 The Data Sets

There are basically three types of cadastral data being considered for future cadastral information systems development at the Land Information Centre. At present our thinking is as follows:

* There is already one cadastral data base product at the Land Information Centre, the **digital cadastral data base** (DCDB). The data is mainly a digital replica of hard copy map products. This product needs to be recognised for its strengths and weakness. It has a client base which is satisfied with digitised graphical data. This needs to be preserved and supported for its existing and future client base.

  This data set contains graphical spatially referenced land information for New South Wales that is stored in digital form. Basically, the data collected contains coordinates and line work relating to cadastral and administrative boundaries. Some line attribution may also be created.

* The second product that is being considered, and is known as the **proposed layer** of the DCDB, or development applications layer. This product will consist of a separate layer in the existing digital cadastral data base product, consisting of development applications. It is anticipated that this information could be put on line so that data dissemination to relevant utility and government organisations can occur remotely through modem access. Some of this data could be survey accurate, but much will not be. Further, this data is likely to change at various stages through the development process.

  The development application information is useful to the decision makers in the land development cycle. At the end of the development process a deposited plan is produced. At this stage the development application could be removed from the proposed layer, relevant reports can be made available, and actions taken.

  The data contained in this set is basically graphics. The data may change during the development process, and so accurate numeric data is not required. These data sets represent a conceptual view of the proposed development site.

* The third product is the **cadastral spatial referencing system**. This proposed product is to contain only survey accurate, survey derived data. It is anticipated that the Board of Surveyors data base will be used to verify the authority of suppliers of data. The cadastral spatial referencing system is not only a data base containing survey related data, but it is a management system for the data and the flow of data.

The cadastral spatial referencing system is not targeting the client base of the digital cadastral data base. The cadastral spatial referencing system is only concerned with survey accurate data, preferably obtained directly from its source. Therefore the main users of the data will be those who utilise surveying services and data.

The main clients are foreseen as:

1. The internal client - Land Information Centre.
2. Surveyors - including cadastral, engineering, etc.
3. Government organisations
4. Utility services

The data contained in this set is survey derived information, to survey standards. The information set, in part, contains much of the information on the face of a deposited plan. But there is much more. The data set introduces the principle of uniquely identified, externally addressable, sequential point numbers and lines. The concept of unique point numbering is the fundamental principle by which land information/geographical information systems will be able to migrate to a system of "vertical integration" of thematically displaced data sets. That is, correlating, or bringing into correspondence, various conceptual "layers", each developed independently and relating to a specific data set content.

The data managed by the cadastral spatial referencing system has been divided into the following basic components:

a. Points
b. Lines
c. Polygons
d. Headers
e. Dimensions

Further, the cadastral spatial referencing system is not an isolated product. It is one of four related survey data bases that make up the Land Information Centre's survey information system (Section 2.3). Therefore, the data content must not only satisfy external clients, but also the internal client.

As can be seen in the preceding discussion there have been a number of changes in direction for the cadastral spatial referencing system since its inception. Associated with these physical changes there have been some changes in the underlying philosophy behind the cadastral spatial referencing system.

Finally, the present system of land information/geographical information systems data management, only allows for spatially registered data. It will not allow parametric relationships to be established between spatially referenced entities. This visionary aspect of the cadastral spatial referencing system is over-looked by all decision makers responsible for this project's progress.

The aims of the cadastral spatial referencing system are;

a. to manage the flow of digital survey information,

b. to manage the spatial aspects of the cadastre,
c. to up date the existing products,
d. to keep the Land Information Centre at the fore front of information systems technology.

3.3 The Data Bases

The present philosophy at the Land Information Centre is that the three products outlined will satisfy a very diverse range of clients.

Firstly, those clients that have purchased the existing digital cadastral data base will, for a significant time still require a cadastral map fabric which has low spatial resolution, is fixed and unmovable, and is only updated every twelve months. Nevertheless, this product suits the needs of many clients.

The data in the digital cadastral data base is basically stored as proprietary files, with some attribution being stored in a relational data base.

Some clients now require the data to be updated more regularly. Indeed, they need to know in advance if a deposited plan is being registered, and that that activity may occur in a particular area. Such organisations need information to be available early in the development process in order that they can plan for infrastructure development. At present the existing administrative system requires that such information be circulated to utilities and authorities in multiple hard copy format. The dissemination of this information can be implemented by the local government agency involved, or by the client, depending on local circumstances.

The provision of a standardised electronic format for disseminating this type of information will lead to efficiencies at the design stage of infrastructure, and therefore lower turn around times for utility planning.

In order to develop this capability the Land Information Centre has committed itself to the development of a proposed layer in parallel with the digital cadastral data base. The Local Government organisations may then encourage data submission in electronic format, which will be added to the proposed layer. For this purpose, data in any proprietary transfer file structure would be adequate (eg DXF).

Some clients require more accurate information, and they require that information to be available and up-to-date at all times. In order to meet these needs, and the needs of the Land Information Centre, the cadastral spatial referencing system has been proposed.

The cadastral spatial referencing system is the only system that proposes a cost effective methodology for progressively populating the digital cadastral data base data set with newly developed cadastral infrastructure. These small bundles of data must find their way into the digital cadastral data base, and the cadastral spatial referencing system can be the facilitating mechanism.

The data for the cadastral spatial referencing system will be stored in a relational data base. The data will be stored in a context independent manner, thereby not relying on any proprietary structure. This will ensure easy portability of the data between platforms.

4. THE FUTURE OF THE SPATIAL CADASTRE

4.1 The Responsibility

The Surveyor General has a responsibility to ensure the integrity of the state's spatial cadastre. With the ever increasing rate of technological advance in information systems, the ability to manage, monitor and improve the state's spatial cadastre is becoming a more realistic objective. Further, the Land Information Centre promotes itself as a world leader in information system technology. Therefore, if it is to maintain that world leadership status it must be at the leading edge of development. The cadastral spatial referencing system will ensure that the Land Information Centre stays in front.

4.2 The Need

Spatial location, both relative and absolute, will become more important to people as technology such as GPS become more affordable and reliable. Further, as computer technology filters down to all aspects of society, people are going to need access to data, more readily, and in different formats. We must be willing to meet this challenge.

4.3 The LIS of the Future

The future of land information systems at the Land Information Centre looks promising, particularly if the cadastral spatial referencing system is implemented.

Data would be gathered directly from its source - the cadastral surveyor. The cadastral surveyor as a collective is the only group that has the imprimatur to create or change the cadastral fabric of the state. This data could be transmitted to the system via modem up load procedures. The data quality will therefore be of the highest order, not having undergone cartographic degradation.

Further, the proposed flow of this information is cost effective. The cost burden of updating the cadastral representation is shifted from the state government, and therefore all tax payers, to those that interfere with the cadastral infrastructure (ie the land developer).

Once the data enters the system it will be managed efficiently. The spatial aspects of the cadastre will be managed, analysed and improved.

Data will also be available on-line in electronic format, or by standard inquiry systems. A simple dial up service will give all account holders efficient access to vital data, when they need it.

The public has paid dearly for the management system that has developed to date. It would be irresponsible not to support or improve the present investment in land information systems at the Land Information Centre.

Further, the future for land information systems will be related to the value that good and timely information will attract. The Land Information Centre will provide significant value to information in the future as it develops appropriate technologies today. The Land Information Centre could become a significant data broker in the future.

5. Benefits of a Cadastral Spatial Referencing System

The next generation of land information systems to be developed at the Land Information Centre will bring many benefits to the spatial aspects of the cadastre. It will;

i. Provide common referencing for all surveys, and thereby improve the spatial aspects of the cadastre,
ii. Provide a framework for cadastral information recovery,
iii. Improve the quality and availability of surveyed cadastral information in New South Wales,
iv. Make available intelligent digital cadastral information.

6. CONCLUSION

Ultimately a decision has to be made as to whether the state of New South Wales wishes to pursue a position of world leadership status in information systems. This will also lead to significant benefits for the state. That decision falls on three groups of people:

i. the New South Wales public must want such a system, and therefore they must be made aware of the efficiencies that can be obtained,
ii. the surveying profession must support the concept, as they are the data suppliers, and would possibly be the main users of the system,
iii. the senior bureaucrats and politicians within the government must understand and then made to support the concept, in order that an efficient development and implementation is obtained.

The Land Information Centre is aware of its responsibilities towards the development of the spatial aspects of the cadastre, and the value of spatial and aspatial data in electronic format. It has determined that it is its "business" to manage the spatial aspects of the cadastre, and is making this decision clear to its clients.

The Land Information Centre needs to seriously consider the following question when putting together a development strategy which involves public participation.

> "Are the perceived clients technologically ready for managed electronic survey information and the benefits resulting from the information flow efficiencies?"

7. BIBLIOGRAPHY

Aptech, 1986, **Feasibility Study into a Coordinated Cadastre for South Australia**, Technical Report 1/86, South Australian Department of Lands.

Angus-Leppan, P.E. and Angus-Leppan, P.V., **1990, Report on Cost Benefit Analysis of "Establishing a Cadastral Coordinated System in the A.C.T.**, Department of Environment, Land and Planning, Canberra, ACT.

Baitch, G. and Pilgrim, L., 1993, **Land Information Systems at the Land Information Centre, - the past, present, and future**, NSW Staff Surveyors Association, 69th Annual Conference, Gerringong.

Baitch, G. and Toms, K.N., 1989, **Planning For a Cadastral Spatial Referencing System For New South Wales**, Technical Report, Land Information Division, Department of Lands (Land Information Centre).

Haddon, K., and Egan, C., 1991, **Coordinated Cadastre System Development**, NSW Staff Surveyors Association, 67th Annual Conference, Bathurst.

Harcombe, P., 1992, **Cadastral Reform in NSW**, NSW Staff Surveyors Association, 68th Annual Conference, Penrith.

Johnstone, T.M. and Toms, K.N., 1989, **Review of Adoption of legal Parcel Coordinated in the Australian Capital Territory**, Technical Report, ACTILIS, Canberra, ACT.

An Approach to Combine Fuzziness with a Physical Model for Landuse Planning Using GIS

Nathaniel C. Bantayan
Department of Surveying and Land Information
The University of Melbourne
Parkville VIC 3052
Australia

and

Ian D. Bishop
Centre for GIS and Modelling
The University of Melbourne
Parkville VIC 3052
Australia

ABSTRACT

Landuse planning and management issues are affected by physical, political and socio-economic factors. Any decisionmaking activity on a piece of land, ideally, should consider these factors in a rational and conciliatory/compromising manner. People's opinions and preferences on these issues can be grouped together to come up with a ranked list of landuse alternatives based on a number of objectives using a fuzzy group multicriteria decisionmaking (MCDM) model. On the physical side, an erosion-based model (EBM) comes up with an index on the degree of erosion for the different landuse alternatives. Using geographic information systems, MCDM is combined with EBM and the results are used to assess the suitability of existing landuses as well as diagnose land capability for specific landuse alternatives.

INTRODUCTION

Landuse planning involves a careful consideration of the demands of a population on the resources of the land. A low density population does not put much pressure on the land while meeting the needs of the people. As population increases, so must the production of commodities to satisfy human needs. While demands increase, land, which is the basic unit of production, remains fixed in available area. The solution, typically, is to increase inputs of labour and energy in order to expand output. Whether this means clearing low energy forests or adding chemicals to formerly low intensity areas, there are inevitable environmental consequences. The ultimate objective of landuse planning, therefore, is to develop sound landuse plans - plans which meet societal objectives while, at the same time, being environmentally sustainable.

Arriving at social acceptance is very tricky mainly due to the fuzziness of human decisions. According to Xiang, et al. (1992), fuzziness has two connotations: (1) it implies human beings' distinct ability to make descriptions and judgments about the behaviour of a system based on fuzzy logic and (2) it represents the imprecision arising from 'fuzzy' description and judgment.

Also, in addition to fuzziness, decisionmaking involves several objectives/attributes (multicriteria) and many individuals (multiparticipants) which encompass multicriteria decisionmaking. Treating these three characteristics (i.e. fuzziness, multicriteria and multiparticipants) of decisionmaking simultaneously have been done by some researchers. Blin and Whinston (1974) modelled group preferences using the theory of fuzzy sets and fuzzy relations. They showed that the opinions of multiple observers can be aggregated to eventually come up with a group consensus. This is useful especially for group decisionmaking where the decisionmaker becomes a collective entity (Blin, 1974). Hipel (1984) applied a fuzzy group multicriteria decisionmaking (MCDM) model to a solid waste management problem in selecting the best alternative solutions. In 1979, Znotinas and Hipel applied a fuzzy set theory to a water resources project. Xiang, et al. (1992), who also cite the works of Fung and Fu (1975) and Sobral, et al. (1981), developed a model in selecting the 'least objectionable solutions' to a landuse planning problem using a fuzzy group MCDM model.

Solutions based on fuzzy theory tend not to include the option of physical modelling. Similarly, physical models commonly allow no room for qualitative variables and subjective assessments. This paper looks towards combining these two valid, but limited, approaches. The fuzzy group MCDM model is integrated with an erosion-based model (EBM) using the analytical and graphical features of geographic information systems. The results of the two models are then used for land capability classification and landuse suitability analysis. On one hand, the fuzzy group MCDM model is treated as a subjective model which considers the opinions of a group of people on a predetermined list of landuse alternatives. A result of this model is a ranking of these alternatives based on identified objectives/attributes which are seen as major factors affecting social choice. On the other hand, a physical erosion-based model generates an index on the degree of erosion for the different landuses.

It must be pointed out that this paper represents an initial stage towards developing a decision support model for landuse planning. The example given here treats each landuse alternative separately via a single-use approach. This research will later be improved to account for the effects of neighbouring landuse alternatives including area requirements.

THE APPROACH

The fuzzy group multicriteria decisionmaking model contains four procedures (Xiang, et al., 1992):

1. Individual Opinions. Establish individual, fuzzy rating matrices that have weightings on all objectives or attributes.

2. Group Opinion. Construct a weighted rating matrix of basic feasible solutions by aggregating the individual matrices using a mixed aggregation rule.

3. Find the 'reduced' rating matrix of noninferior solutions by eliminating non-essential objectives or attributes and inferior solutions from the weighted, aggregated rating matrix.

4. Determine the 'best' or 'least objectionable solution.'

In their model, Xiang, et al. (1992) first defines the alternative scenarios, objectives and attributes, the preference measurement system and the interested individuals and groups. This is followed by data collection via a questionnaire asking the respondents to judge the alternative scenarios based on the objectives and attributes using an appropriate preference measurement system. For computer processing, the 'fuzzy labels' are transformed into 'evaluation values.' For example, for grades of membership (Sg), the 'fuzzy labels' very high, high, moderate, low and very low correspond to the 'evaluation values' 1.0, 0.7, 0.5, 0.3 and 0.0, respectively.

A physical model presents the advantage of measurability. In this paper, a physical model based on soil erosion is chosen because it is the single, most useful indicator of land capability (Cruz and Ffolliott, 1991). It is mostly affected by factors such as topography, slope, existing landuse, vegetation cover and rainfall. These factors can be spatially referenced and geographic information systems can be used to advantage.

These two models, used separately, can assist decisionmakers in landuse planning and management. The subjective model provides information on group consensus regarding landuse alternatives. The physical model provides empirical information on the degrees of erosion for specific land management units. Taken together, however, they can vastly improve the decisionmaker's options and thereby assist in making better decisions.

Geographic information systems provide the link between the two models as shown in Figure 1. The inputs to MCDM are individual opinions while the inputs to EBM are the bio-physical parameters

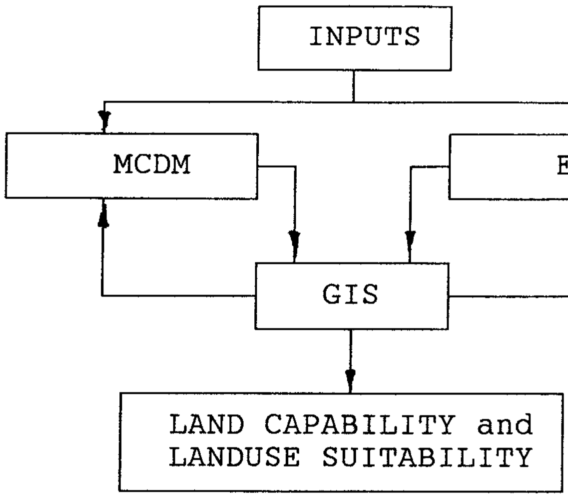


Figure 1. Conceptual framework of the model.

(e.g. climate, topography, slope, existing landuse, vegetation cover). The results of the two, in this case, submodels, are combined within the framework of GIS. However, the difficulty in combination lies in the aspatial nature of the MCDM results. In other words, the 'least objectionable solutions' indicate only the extent to which one landuse alternative is preferred over another or the equal acceptability of a number of landuse options. Work reviewed to date will not local the preferred landuses on the ground. The preferred uses still have to be spatially allocated. This will involve combination of the preference relationships with data about physical, environmental and social character of the land in question. This data may be original mappings or the output of models dealing with habitat, productivity, recreation or other resources, in addition to erosion liability.

THE CASE STUDY

Melbourne Water is preparing a masterplan for Plenty Gorge Park with the goal of providing for the long-term protection of its landscape and environment, as well as a source of recreation for northern Melbourne. In support of this masterplan, a GIS study was conducted to assess several issues affecting the Park. Some data from this GIS study was used for this particular case study.

This case study made use of the following landuse alternatives: open land, forested land, active recreation, abandoned quarries, active quarries, low density buildings, high density buildings and roads. For the erosion-based model, the objective was to determine the erosion potentials of the landuse alternatives. To arrive at these potentials, soil erosion was assumed to be affected by soil characteristics, slope, landuse, vegetation cover and rainfall. These factors were generated as coverages using a raster-based GIS package called IMAP (Itami, 1992). An erodibility coverage was generated by a combination of the soil and slope maps based on the assumption that soil erodes more at higher slopes. Land cover was represented by an overlay procedure between the landuse and vegetation cover maps. A uniform rainfall was assumed on the whole area. These coverages were further combined to finally produce an erosion coverage showing a nominal erosion index (i.e. low, moderate, high and very high erosion hazard).

Table 1 shows the percent degree of erosion for each landuse alternative. Comparatively, forested areas have the least erosion potential while active quarries have the highest. For the fuzzy group MCDM model, the same landuse alternatives were used and evaluated against several objectives. These were: maintenance and enhancement of habitat values, connectivity and quality; susceptibility to fire; contribution to soil erosion and productivity*; compatibility of landuse alternatives adjacent to the Park; contribution to visual impact on the Park; contribution to the general public access to the Park; and contribution to overall environmental quality.

Landuse alternatives\ Degree of erosion	low	moderate	high	very high
open land	62	22	13	3
forested land	78	14	6	2
active recreation	0	16	76	8
abandoned quarries	0	0	80	20
active quarries	0	0	40	60
low density bldgs	31	38	19	12
high density bldgs	57	26	15	2
roads	0	59	23	18

Table 1. Percent degree of erosion per landuse.

Respondents to the questionnaire were asked to use these objectives as evaluation criteria for the landuse alternatives. A preference measurement system following a continuous interval and linear order was used (i.e. 0.0, 0.3, 0.5, 0.7 and 1.0) for the grades of membership as well as the weightings.

Table 2 shows the set of overall grades of membership in a ranked order. As shown, a mixture of landuse alternatives receive the same ranks. It is apparent, however, that a decisionmaker would

----

*inclusion of erosion among the subjectively rated objectives amounts to double counting of this criterion. Objectives which can be physically modelled should be treated separately in the early stages of the process.

not authorize active quarrying within the area, firstly because the people do not want it and secondly due to the damage it causes to the landscape (see Table 1). The highest ranked alternatives (rank 1) are also those found to pose the least erosion hazard except for the abandoned quarries. It is also interesting to note that while the Park is a recreation area, active recreation received a moderate ranking.

Rank	1	2	3
Landuse alternatives	-open land -forested land -abandoned quarries -low density bldgs	-active recr. -high density bldgs -roads	-active quarries

Table 2. Ranking of landuse alternatives.

DISCUSSION

While the case study provides some initial insights, it still only partially uses the capabilities of a geographic information system as the basis for integrated physico-subjective modelling. A number of significant developments are planned for further integration of the system components identified here.

At this stage, for example, each landuse alternative is treated separately without consideration of the effects of neighbouring landuses. In the real world, a multiple use and locationally interdependent approach to landuse planning is inevitable. To accommodate this requirement, it is intended to implement the decision support system on a cell-by-cell basis. Each cell will have attributes derived from both physical models and subjective assessments. These will together contribute to suitability for particular landuse alternatives. Once the MCDM and the physical models are working from the same data sets and writing results back into the same data environment (the GIS) the basic information required for an allocation operation will be in place.

This allocation phase will also require a mixture of scientifically derived and subjective inputs. For example, a minimum or a maximum area for a specific landuse option may need to be defined. In the case of habitat requirements for a particular species this may be quantified by the scientific community. In the case of recreational facilities, these ranges can come from a subjective-fuzzy-procedure.

Dealing with the eventual combination of space requirements, suitabilities and preferred spatial interrelationships is a complex allocational exercise. We foreshadow an iterative computational solution based on accretion of landuses around seed cells allocated either on the basis of clear suitability superiority or by direct designation by a decisionmaker.

Thus, at all stages of the process of landuse allocation, there will be synthesis of direct and fuzzy procedures. The entire system will eventually be tested in a multiparticipant, multicriteria workshop in Los Banos, Philippines.

REFERENCES

Blin, J.M. and Whinston, A.B. (1974), "Fuzzy Sets and Social Choice", Journal of Cybernetics, Vol.3,No.4, 28-36.

Blin, J.M. (1974), "Fuzzy relations in Group Decision Theory", Journal of Cybernetics, Vol.4,No.2,17-22.

Cruz, R.V.O. and Ffolliott, P.F. (1991), "A GIS-based Land Capability Classification and Land Use Suitability Assessment of Ibulao Watershed, Philippines", PTEROCARPUS A Philippine Science Journal of Forestry and Forestry Products Vol.6, No.1, March 23-32.

Hipel,K.W. (1982), "Fuzzy Set Methodologies in Multicriteria Modelling", Fuzzy Information and Decision Processes (MM Gupta and E.Sanchez, eds.), North-Holland Publishing Co., 279-287.

Itami, R.M. (1992) "IMAP reference manual" School of Environmental Planning, The University of Melbourne, Australia.

Xiang, W-N, Gross, M., Gy Fabos, J. and MacDougall, E.B. (1992), "A Fuzzy-Group Multicriteria Decisionmaking Model and Its Application to Land-use Planning", Environment and Planning B: Planning and Design, Vol.19,61-84.

Znotinas, N.M. and Hipel, K.W. (1979), "Comparison of Alternative Engineering Designs," Water Resources Bulletin, Vol.15,No.1,44-59

INFORMATION-ORIENTED BUSINESS

John W. Blackburn
Genasys II Pty Ltd
Level 13, 33 Berry Street
North Sydney 2060
Australia

ABSTRACT

The aim of this paper is to describe the relationship between Land Information Management principles and methods, and "information-oriented business". An example is used to exhibit the transition from being automated to being information-oriented. The focus is not on the end-user and technology functionality, it is on the corporate mission, business goals and customers. Technology, in particular GIS, is presented as the enabler, how it has been applied and what trends are seen in its development in light of LIM. Data availability is described as the facilitator.

INTRODUCTION

Information oriented businesses are those which have made the transition from using computers to mechanise standard business practices such as accounting, word processing and inventory control to using computers to integrate data into information to support decision making. Whether systems are centralised or decentralised fragmentation of data and data protectionism serve to maintain a mechanistic approach i.e. locking data and not permitting it to flow in the organisation. More recently using data as a resource has led to more productive use of computing to permit modelling through the combination of data and generation of information. The value of integrated information is not just to senior management it is an integral part of the decision support, performance monitoring and process control from board room to shop floor and over-the-counter service.

A further stimulus to the more imaginative use of data is through the incorporation of location data which provides information with a spatial expression. The use of land information FOR management has provided an additional impetus for data integration which requires:

- multi-disciplinary data and expertise;
- inter-related spatial and descriptive data.

A Geographical Information Systems (GIS) is an essential component, in fact an enabling technology for using land information for management. A successful implementation necessitates that the system be permitted to transcend professional, organisational and operational boundaries. The challenge is therefore more often procedural and cultural than technological. By addressing procedural issues within a framework which is focused on data integration, companies and organisations are becoming information-oriented and opening many new opportunities to enhance production, service and profitability.

INFORMATION-ORIENTED - A PERSPECTIVE

An information-oriented enterprise is a company (the term "company" is used generically) in which information is fundamental to any decision making process. Information oriented companies often develop from "automated" organisations. In automated organisations, computers primarily are used to automate labour-intensive and repetitive operations. In information-oriented companies, information technology is strategically aligned with corporate objectives, encompassing a wide spectrum of convergent and linked technologies.

A Perspective on Location

It is hard to think of many business determinants or variables, at least at the local level which are not location-related. The relationships to business performance are often direct, e.g. sales and turnover in the case of a service centre, can be directly related to access, volume and type of the customer catchment, proximity of surrounding ancillary or complimentary business activities e.g. supermarkets, varied retail outlets and merchandising. Locational determinants need to be analysed in context and each to have a certain favourable rating or weighting. The right mix is important, too much of one could cancel out the beneficial effect of the others; for example, siting at a major shopping mall may cancel out the effect of surrounding business because of restricted access from major roads.

A Role for GIS

Geographic Information Systems (GIS) are designed specifically for dealing with "location", maybe the technology should be called LIS i.e. Locational Information System. Either term is somewhat limiting as the technology can play a wider role in providing information solutions which integrate data and information from a variety of applications and databases having both internal and external sources. However, the common thread or integrating mechanism is the locational identifier e.g. address, co-ordinate position values etc; and the common requirement is to visualise quantitative data and analyse it spatially (locationally).

Improving the efficiency and effectiveness of existing business activities is an important objective particularly in changing and competitive economic conditions. There is a role for GIS technologies in supporting these initiatives. In this role GIS may need to be established in several operational areas to integrate data and be used selectively, closely aligning it with existing operational practices. To realise early benefits of GIS there needs to be a strong commitment from management and acceptance from operational areas. Inter-twining GIS technologies and techniques into the company's operational activities and aligning them with corporate information management strategies should be a planned and gradual process. This may involve complex issues that require a decision support system with a geographic context, as well as the ability to develop several decision models that can correlate numerous geographic attributes for integrated solutions. GIS can be a decision support system.

BUSINESS DEVELOPMENT USING GIS IN THE 90's

Geographic Information Systems have undergone a significant transformation in the past 5 years. GIS as both a technology and methodology has matured from computer-assisted mapping into an information system for "spatial decision support and analysis".

The availability and coverage of spatial or geographic databases has also increased considerably covering a wide range of themes and uses such as engineering, infrastructure, land use, and demographic data which are available from government agencies and utilities in digital form. LIM has facilitated this situation via various government co-ordinating initiatives.

What is fuelling the application of GIS techniques is the realisation that a great proportion of a company's business, customers or assets are spatially related or dispersed and that GIS offers integrated applications to solve business problems and provide new marketing tools.

Most companies and organisations now use databases to store data, but most importantly to access and analyse information, Information systems such as EIS, MIS and now GIS can work together to share and integrate information because of the common architecture i.e. "the database".

GIS is a "database" which has a graphic user interface to visualise the database and a suite of spatial analysis tools to produce new information.

AN EXAMPLE OF BEING INFORMATION ORIENTED

A good example of information-oriented business is in Local Government, arguably the richest area for land information utilisation and applications. There are many examples of applications of GIS technology and LIM principles in Local Government which provide practical benefit to Council operations and the community.

The Local Environment Plan (LEP) in NSW, is the responsibility of the local council under the Environmental Planning and Assessment Act, 1979. The LEP supports land use policy and development control, and by virtue of this provides a framework for the LIM "integrating" approach. The LEP is more than the integration of various data layers, such as Cadastral Property Data, Zoning, Land Ownership, Infrastructure and Demographic data.

The LEP exemplifies how data is built into information which is interpreted in light of public interest and presented as policy to guide officials in decisions concerning land use. GIS provides the enabling technology.

The Implementation Strategy

Genasys has been involved in Municipal Management for more than 15 years. It has witnessed and participated in the introduction and acceptance of GIS as part of the LIM approach.

Local Government is principally about "land management" where trends from "automation" as an initial need, have developed into "information-orientation". Genasys is seeing implementation strategies being applied which exhibit these characteristics. The following strategy is representative (viz: Orange City Council GIS Implementation Strategy, January 1993), where business objectives are described in a plan, incorporating:

- Goals
- Inputs
- Resources
- Organisation
- Priorities
- Outputs
- Monitoring

The plan is and must be simple, have defined milestones and deliverables, and must be defensible for accountability.

A perspective and conceptual view for management is provided at Diagram 1.

A central goal would be:

> "create a new Town Plan, to improve operational and decision making performance through a focus on locational and spatial analysis".

Tasks and functions as applications are defined in terms of their minimum data requirements, (summarised in Table 1), and typically a textual land information database will exist or is a prerequisite which incorporates property, rating, financial, and records sub-systems.

Resources, organisation, outputs and monitoring are not included in this discussion for reasons of brevity, even though they are key aspects of the plan.

These are by no means the only applications available within the GIS. In fact the list of potential applications or uses is almost endless, and is limited only by the data available to be accessed. Some other applications for use in each division would be:

APPLICATION	UTILISATION	INPUTS REQUIRED	DATA AVAILABILITY
LOCALITY RATING	Rates; Financial Planning	Rates System	Existing and as provided by VG
ADJOINING OWNERS	All areas where public notification is required.	Property system; Standard letter generation	In-built in System
ZONING	All property system users	Zone codes	Data input to/from Property system
UTILITIES	Infrastructure Planning; Development; Building Assessment; Maintenance.	Utility services survey and input	Survey results input, and conversion from engineering drawings
DRAINAGE DIAGRAMS	Community Service; Development Assessment;	Property surveys	From site surveys
DEVELOPMENT TRACKING	Strategic Planning; Personnel Management; Service Delivery	Property system applications e.g. Building and Development (BA/DA)	Data entry from on-line applications
COMPLAINT TRACKING	Maintenance programming; Resource management; Service Delivery	Complaints procedure (Records Management)	Data input as part of complaint registration system
COMMUNITY SERVICE DELIVERY	Management; Service Delivery	Link to Community Services System	Health care program Education program
ROAD SERVICE	Traffic planning; Maintenance programming; Pavement Management;	Traffic flow info. Pavement Management system link	Road maintenance budget; Digital road network
CENSUS APPLICATIONS	Planning for provision of services	Demographic statistics; Collection District mapping	Community Profiles from ABS;

Table 1 : Relevant Applications

Corporate services;

Crown land, Council land, public land, vacant land, church properties, park land, roads, street addresses, ownership details, rating details, etc.

Environmental services

Heritage orders, erosion, parking, open space, building outlines, dog licenses, garbage services, catchment management, fire notices, health notices, weed notices, etc.

Technical services

Contour generation, water catchments, terrain modelling, road planning, pavement management, bus routes, street lighting, survey marks, assets, playgrounds, park and street furniture, utility buildings, dams, etc.

Human services

Location of public buildings, tracking of service recipients, providers and venues (for aged, daycare, before and after school care), public building floor plans (library), population statistics by areas, etc.

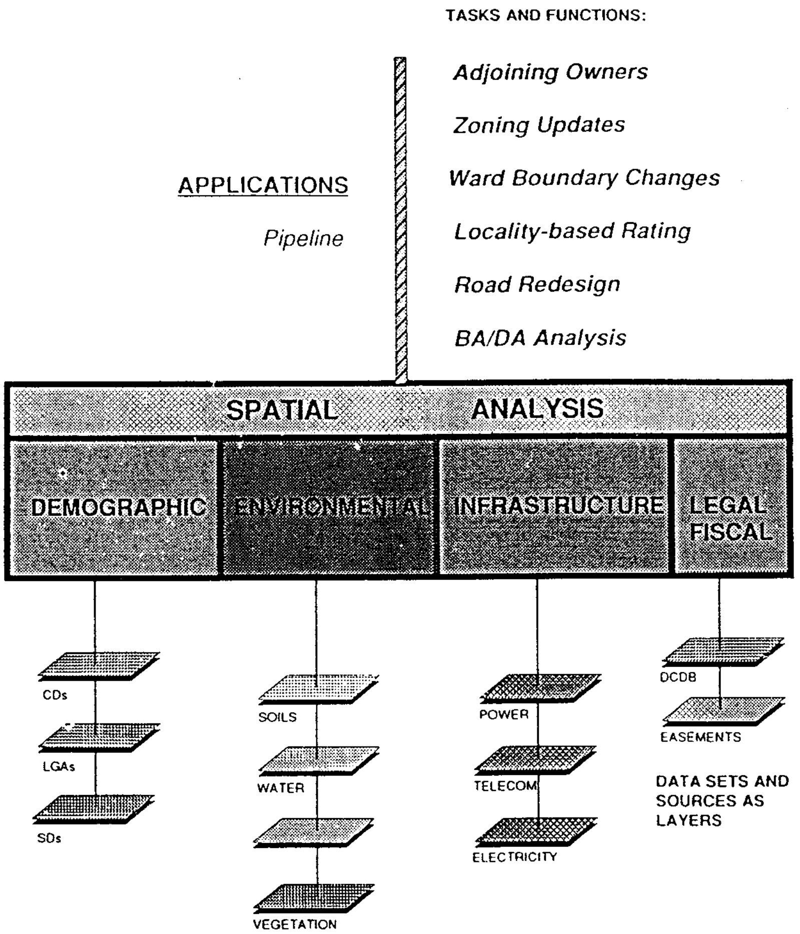


Diagram 1: Management Perspective of the Geographic Information System

Priority Applications

Local Rating (Priority 1)

This application has identifiable benefits to the Corporate Organisation since the Locality rating system is fundamental to the rating and financial structure of Council and importantly provides a fairer system of rating to ratepayers generally.

For locality rating to be represented in the GIS, the land value and "Rate Code" fields from the Rates sub-system need to be related to land parcels in the GIS. To do this requires cadastral information in the GIS and the rates/property systems to be matched.

The use of GIS for the revision of the Locality Rating System is able to show considerable productivity gains.

Adjoining Owners (Priority 2)

The GIS provides for identification of properties that may be affected by development or action on adjoining land. The system involves identifying the land and setting criteria to search surrounding properties.

Examples of information provision:

- Notification of persons considered to be affected by a development proposal (adjoining, adjacent and in the vicinity).
- Notification of owners (adjoining only) advising that a building application has been received.
- Notification of residents (adjoining) along a stretch of road or utility that an interruption to service is to occur.
- Notification of residents (area) regarding a change to garbage services.

The use of GIS allows the visualisation of affected properties, this is favoured since neighbouring owners may not simply be those with sequential street addresses. Once the GIS/property data is verified and the property system is active the "adjoining owners" application is able to be used. The only digital entry required is the creation of standard letters linked to the system.

There would be virtually no cost involved in providing this application. The benefits include time saved in determining adjoining owners and more accurate identification and therefore improved customer service. The adjoining owner application utilising GIS provides quick, accurate notices to adjoining owners. This application has capabilities for many areas of Council. It therefore meets the priority criteria for utilisation of the GIS. It is a clear productivity tool.

Zoning (Priority 3)

The creation of a digital zone map (LEP No. 11) is a recommended application. The property system has a strong emphasis towards zoning information with its links to Section 149 Certificates. The property system is also built up on zone codes which have been created.

Since the L. Shaddock P/L v Parramatta City Council finding in 1981 where a Council was required to pay $ 173,938 plus court costs for negligence in relation to planning advice and a zoning certificate, it is imperative that accurate up-to-date zoning information is available.

Because of the number of changes to the LEP a Master Map is required to be kept under the Environmental Planning and Assessment Act. There is typically only one master map currently held in Council. The creation of the LEP in the GIS provides for a master map to be held and available in a number of places simultaneously. The GIS provides a cross check with the property system to ensure that zoning information is accurate. The GIS map will be built up from the property system and checked against the LEP. Zoning information will be inserted into the property system and then loaded across to the GIS. The system relies on the input of zoning codes into the property system and parcel data in the GIS.

The Catchment boundaries are created to define the ridge lines and extent of water quality control areas. The distributor road boundaries together with the catchment boundaries provide the most common examples of "split" zones.

The Section 149 Certificate system is reliant upon the property/GIS system. Section 149 Certificates are a significant source of revenue. The automation of the Section 149 Certificate process can reduce processing time from an estimate of 30 minutes for a basic Certificate (requiring involvement of three to four people) to 5 minutes per certificate with automation requiring involvement of one to two people. The average turnaround time for a certificate application is currently 4.5 days. This time can be expected to reduce to an average 2 day turnaround period through the application of the GIS/Property system.

The creation of a zone map in the GIS satisfies the priority criteria since it provides a benefit to the organisation through ensuring accurate property information to a number of users simultaneously; achieves a productivity increase in accurate certificates; and relies on the data already in the GIS and information required for the property system.

Collation of Data

Now in the 90's there is a clearer understanding of the data requirements and their availability, cost etc. for these applications. De-mystifying and simplifying data acquisition and integration has been a key part of this, and is exemplified by the following process:

Priorities

Must have:	• Cadastre (DCDB)
	• Link to Property information (includes matching and "clean-up" process)
	• Zoning and other textual property information
	• Knowledge of data issues (e.g. accuracy)
Should have:	• Contours
	• Road network
	• Aerial photographs (for backdrops)
Other:	• Soils
	• Vegetation
	• Telecom
	• Water/Sewerage
	• Drainage

Therefore, get the "must haves" and you have the raw materials for the 3 priority applications described.

From Operations to Decision Support

The "adjoining owners" application is a demonstrable productivity improver in the day-to-day business operations. Its logical extension is forecasting and monitoring. For example, monitoring of Building Applications (BA) and Development Applications (DA) activity using the GIS can assist in strategic planning, service programming and resource management. Areas of high activity can be identified and resources allocated. Services can be planned when activity reaches certain thresholds and decisions can be made on planned release of land as development areas reach capacity.

This information then feeds back into the LEP and the loop is completed. However, the same information (ie from the LEP) can permeate into the State Government level e.g. into SEPP (State Environment Planning Policy) and further into Commonwealth policies and schemes for environmental monitoring and protection e.g. water quality.

TECHNOLOGY AS THE ENABLER

In the LIM approach, GIS can be regarded now as a premier enabling technology. There are now visible trends in GIS which focus on the issues of integration of systems, methods and data.

Systems

Systems which were once disparate, are now being interfaced, networked and integrated. GIS are now linked to Document Image Management, Road and Pavement Management, Engineering and various other Database systems.

GIS are also being driven by these systems, for example:

- Database driven - records, property etc;
- Analysis-driven - network modelling, statistical.

Methods

Experience, knowledge and logic gained through earlier application of GIS and even automated-mapping is being used in developing:

- Rules-based applications for;
  - intelligent text-placement,
  - cartographic feature arrangement and symbolism,
  - vertical topology i.e. feature-to-feature relationships,
  - automatic road-centreline generation, and
  - network analysis.
- Graphical User Interfaces (GUI) which;
  - are intuitive,
  - provide application executives,
  - employ graphic thresholding,
  - are iconic, and
  - are process as well as command driven.
- Geo-processing functions which are;
  - context sensitive,
  - super-sets of atomic commands to build new commands,
  - process and event-driven,
  - support consequential analysis, and
  - allow embedding of user-defined algorithms.

Data

Multi-custodian and multi-disciplinary databases are now becoming available for access, allowing:

- multi-data type (raster, vector, image, text) integration technologies to be employed,
- employment of "client/server" technologies for distributed, networked databases,
- application and data packaging,
- employment of self-describing database schemes,
- meta databases - data about the data, viz; enough information to know whether the data is useful to the business application,
- temporal data models to manage and preserve historical information viz; "Archival GIS",
- employment of new update regimes and techniques i.e. complete, block or incremental,
- use of mass storage technologies, such as optical storage and data compression schemas,
- security schemes to protect investments in valuable databases.

DATA AS A FACILITATOR

Data availability has long been seen as the facilitator for LIM technologies such as GIS. There have been many initiatives towards co-ordinating the collection, storage and dissemination of land information at State and Federal levels of government. A directory approach may have been taken viz: LANDSEARCH, FINDAR; or a data distribution and access approach viz: ERIN, SLIC.

There are now many maturing land information databases which are providing data to data consumers which is at trivial and premium cost; in a raw, quality assured or application ready state; or is still coveted, owned or restricted.

Data barons are emerging, gearing up to meet demand and embracing the communications revolution which will support the "data highways" of the future.

Genasys has witnessed a clear trend in recent years where most new GIS installations have a corporate database of aspatial data requiring integration with externally available spatial data and virtually no digitising will be necessary by the organisation.

CONCLUSION

In summary, in an information-oriented business, a generic view of "information requirements of corporate database management systems" is:

Role	Database System
Strategic Planning	EIS Executive Information System
Management	MIS Management Information System
Operations	Job Scheduling Assets Register Works/Order System
Visualisation	GIS and Graphics
Analysis	GIS and Modelling Systems

LIS/GIS in the 90's are no longer the domain of special interest groups such as planners and engineers functioning as techniques for data collection and integration. They are an integral component of decision support, management and policy formulation. They provide records of change and as such accountability in the private and public sector.

REFERENCES

Smith, J., Wannan, A., Williamson, J., (1993), "Geographic Information System Implementation Strategy", Orange City Council, January.

Blackburn, J. W., Phillips, M. (1989), "Geographic Information Systems - Defence Applications", Brassey's Australia, Ch10, 204-231.

Blackburn, J. W., (1989), "An Integrated GIS for Planning and Environment", LIS in Local Government Seminar, Melbourne, June.

Blackburn, J.W., Williamson, I.P., (1985), "Land Information Systems in Australasia - Lessons and Experiences", URPIS 13 Conference, Adelaide, November.

PRACTICAL AND PHILOSOPHICAL ASPECTS OF ENVIRONMENTAL MODELLING WITH GIS.

P.A. Burrough,
Department of Physical Geography,
Institute of Geographical Sciences,
University of Utrecht,
Post box 80.115,
3508 TC Utrecht,
the Netherlands.

ABSTRACT

Linking models of environmental processes to databases in a GIS is technically easy, but currently there are few opportunities to evaluate the quality of the results beyond the beauty of the cartographic display. Model errors include both conceptual shortcomings and imprecise specification of coefficients. Data errors relate to conceptual units, spatial and temporal variability, measurement method and noise. Modern GIS should permit users to examine how errors in data and models contribute to errors in results, and how error levels are related to costs. A practical example of predicting the zinc content of river floodplain soils demonstrates the use of new error propagation software (ADAM) and geostatistics for making a comparative cost-benefit evaluation of alternative methods.

INTRODUCTION

There is no more common error than to assume that, because prolonged and accurate mathematical calculations have been made, the application of the result to some fact of nature is absolutely certain.
A.N. Whitehead. (Science and the Modern World. CUP 1953)

Science ... is ..the search for algorithmic compressions. We list sequences of observed data. We try to formulate algorithms that compactly represent the information content of these sequences. Then we test the correctness of our hypothetical abbreviations by using them to predict the next terms in the string. These predictions can then be compared with the future direction of the data sequence. Without the development of algorithmic compressions of data all science would be replaced by mindless stamp collecting - the indiscriminant accumulation of every available fact. Science is predicated upon the belief that the Universe is algorithmically compressible... a belief that there is an abbreviated representation of the logic behind the Universe's properties that can be written down in finite form by human beings.
J.D. Barrow. (Theories of Everything, Vintage, 1991, p.11)

Geographical information systems (GIS) are being increasingly used in government, business and private activities ranging from navigation for ships, planes and automobiles, the registration and organization of pipeline, gas, water, telephone and electricity networks, and for city and regional planning and administration. They are also used for inventory, analysis, modelling and decision making in natural resources studies, agriculture and forestry, and this group of applications will be the main subject of this paper.

By the term "Geographical information systems" I mean the set of hardware and software tools that enable the storage, retrieval, analysis and presentation of spatial data (Burrough 1986), rather than the whole complex of organization, personnel and data that some define as a GIS (Maguire et al, 1991). For the analysis and modelling of environmental phenomena, the GIS tools must be provided with data and with sensible analytical models. If these are in place then the data can be accessed, transformed and manipulated interactively to study environmental processes or to analyse the effects of trends, or for anticipating the possible results of planning decisions. By using the GIS as a trainee pilot would use a flight simulator, it is possible for planners and decision makers to explore a range of possible scenarios and to obtain an idea of the consequences of a course of action before the mistakes have been made irrevocably in the landscape itself.

In this context, GIS can link understanding in the form of *models* to data to explore how the world behaves to achieve new *insights* or to impose some form of *control*. Increasingly the ability to model environmental processes is not limited by computer technology but by human understanding, encapsulated in mathematical models, and the availability of sufficient reliable data to drive them. So, the added value of using GIS for environmental modelling can be expressed as

$$\textit{information} = \textit{conceptual models} + \textit{data}$$

This is only satisfactory if the *results* are not misleading, that is to say that they have been achieved with a certain minimum level of reliability. Just as in any other area of science or industrial production system, the output of a process, whether in the form of useful hypotheses, automobiles, planes, household furniture, or compiled information, needs to meet minimum standards of quality in order to be satisfactory.

Now model building, data inventory and the building of geographical information systems are three separate kinds of activities that until recently have had little to do with each other. Only recently have people found it expedient to move from simple, or complex lumped models of a phenomenon such as runoff or erosion to distributed models that attempt to describe the transport of material over and through whole landscapes. Although model makers have spent much effort on working out the detailed physics and chemistry of processes such as the movement and degradation of pesticides in soil, they have not been specialists in describing how the soil properties critical for their process actually vary, and they often have little understanding of the scale problems involved. Conventional resource surveyors, on the other hand, have frequently recorded the properties of the landscape without recourse to any idea of gradual change, either in space, or as a result of transport processes, through time as well. Linking non-spatial lumped models to static, soil or land use units in a choropleth map is easy in most commercial GIS, but it is a method that has been forced upon us by the glitter of the tools and by a lack of appreciation of the intrinsic aspects of how processes actually operate in a 4-dimensional landscape. An absolutely critical aspect of the whole business of linking models and data in GIS is that there are very few opportunities for determining whether the results are realistic or not. Often the only criterion of "quality" is the cartographic display, which as we all know is of paramount importance when trying to convince someone to invest in GIS.

If *information* = *conceptual models* + *data*, then:

$$\textit{the quality of information} = f\ (\textit{models, data})$$

The quality of models is determined by the degree of understanding of the real world that they encapsulate. A model, be it a logical "IF THEN ...." statement, an empirical regression equation, or a mathematical formulation of a physical process, is in principle, a testable hypothesis. It is also an algorithmic compression of observations that yields a recognizable pattern or recipe from which the original observations can be regenerated. The degree of success is determined by a) the tractability of the phenomenon with respect to algorithmic compression, and b) by our skill in finding a suitable formulation for that compression. Obviously, the more complex the phenomenon, and the larger the number of spatial and temporal scales at which it operates, the less easy it is to find a single, general algorithm that can serve as a "theory of everything".

In the first instance, the quality of data is determined by the conceptual models used to describe the phenomena being observed and the way this information is translated into useable entities (Burrough 1992a). Other constraints on data quality are measurement technique, local spatial and temporal variation, and the physical size of the sample analysed (the *support* in geostatistical terms). Clearly the data entities must be able to support the functions demanded by a given model, for if they do not, then the results will be meaningless. Volume of data is also important, but data quality is not governed by volume alone. Redundancy, spatial and temporal covariation and noise are all important factors that need to be considered, especially when data collection is very often the most expensive part of any environmental modelling operation, once the GIS is in place.

For many GIS users the practical problems concerning the intrinsic quality of results are:

1. How can they identify the levels of quality needed?
2. How can they obtain information about levels of uncertainties in models and data?
3. How can they choose the appropriate method for analysing how errors accrue in GIS modelling?
4. How can they present information about the quality-preserving or quality-reducing aspects of GIS modelling and analysis to a user?

AN ANALYSIS OF THE ISSUES SURROUNDING ERROR PROPAGATION.

Why is error in models with GIS an issue?

If model results are never tested, then there are few grounds for disbelieving predictions made in good faith by serious experts. Failure to test hypotheses, however, is totally unscientific, and it is no excuse to say that model validation is more difficult than model calibration. Moreover, there is a practical aspect in that one expects more complex, "more scientific" models to perform better than simple models. There is a price to pay, however, namely that more complex models usually require many more data at finer levels of spatial and temporal resolution. So both from a scientific and a practical point of view we need to know whether the results are worth the effort. People are reluctant to report that their efforts at more complex modelling do not produce better results, but a recent study by de Roo (1993) concludes that for erosion and runoff modelling "Despite the theoretical advantages of physically-based distributed models like ANSWERS and KINEROS, such as providing hydrographs, soil erosion maps, the effects of conservation planning strategies, testing of scientific knowledge, and indicating future research needs, the quantitative results of those models are not significantly better than the results of simple soil erosion models, such as the USLE and MMF" (de Roo 1993, p. 249). Table 1. summarizes some of his results.

Table 1 Spearman rank correlations (r) of the simulated soil erosion rates for the three storms in the Etzenrade catchment using four models (comparison of the individual ^{137}Cs sampling points and the calculated soil erosion estimates using the Martz & De Jong equation)

3 storms	USLE	MMF	KINEROS	ANSWERS(nc)	ANSWERS(cr)
USLE	--				
MMF	0.927	--			
KINEROS	0.524	0.581	--		
ANSWERS (no crust)	-0.669	-0.747	-0.377		--
ANSWERS (crusted)	-0.692	-0.756	-0.376	0.963	--
137Cs	-0.250	-0.226	-0.158	0.173	0.205
137Cs-est.soil loss	-0.181	-0.138	-0.109	0.113	0.141

0.292 = Significant at 95%
0.150 = Significant at 90%

Determining the level of quality required.

In many cases, GIS users who undertake environmental modelling and analysis have no real idea of the required quality of the results. Research scientists want to understand a process as completely as possible, whereas practictioners and decision makers require results that are sufficiently solid to support a political, planning or management decision. In some cases, such as the levels of toxic chemicals in air, soil or water, strict norms have been defined by health agencies, but frequently it is not known to what level of tolerance these levels need to be identified and mapped. Ideally, all GIS results should be accompanied by error maps that give the associated levels of uncertainty for every prediction.

Levels of errors in models and data.

Models contain two sources of error. The first source is embedded in their conceptual structure - and has to do with the way they make an algorithmic compression of a natural phenomenon. Except for logical models which are discontinuous, most models are linear, because linear modelling is mathematically easy, but many natural phenomena are continuous and non-linear (cf. Stewart, 1989). The second source of error in models has to do with their coefficients. Many environmental models are published without any reference to the standard errors of their coefficients, to the level of residual error, nor to the numbers of data on which they are based. Model coefficients and calibration data are often, for want of better information, assumed to be "error free".

GIS cannot work without data, and data collection is expensive, being often more than 50% of any project costs. Consequently data suppliers are of great importance. There is often discussion about the need to "discover" data, just like discovering oil, so that the GIS engines can be fed. But GIS, and GIS models, are fussy feeders. Not all data will do, and there are many sources of error in data (c.f. Burrough 1986, chapter 6) including age, relevance, level of resolution, size of support (spatial size of the sample), numbers and density that can seriously affect the quality of the results of environmental modelling with GIS. If the quality of GIS results is a function of the quality of models and data, then the *error* in the results of modelling with GIS is a function of the *errors in the models* and the *errors in the data.*

How to choose appropriate methods for analysing error propagation in GIS modelling.

A recent paper (Burrough 1992b) proposed that future GIS should be equipped with sufficient

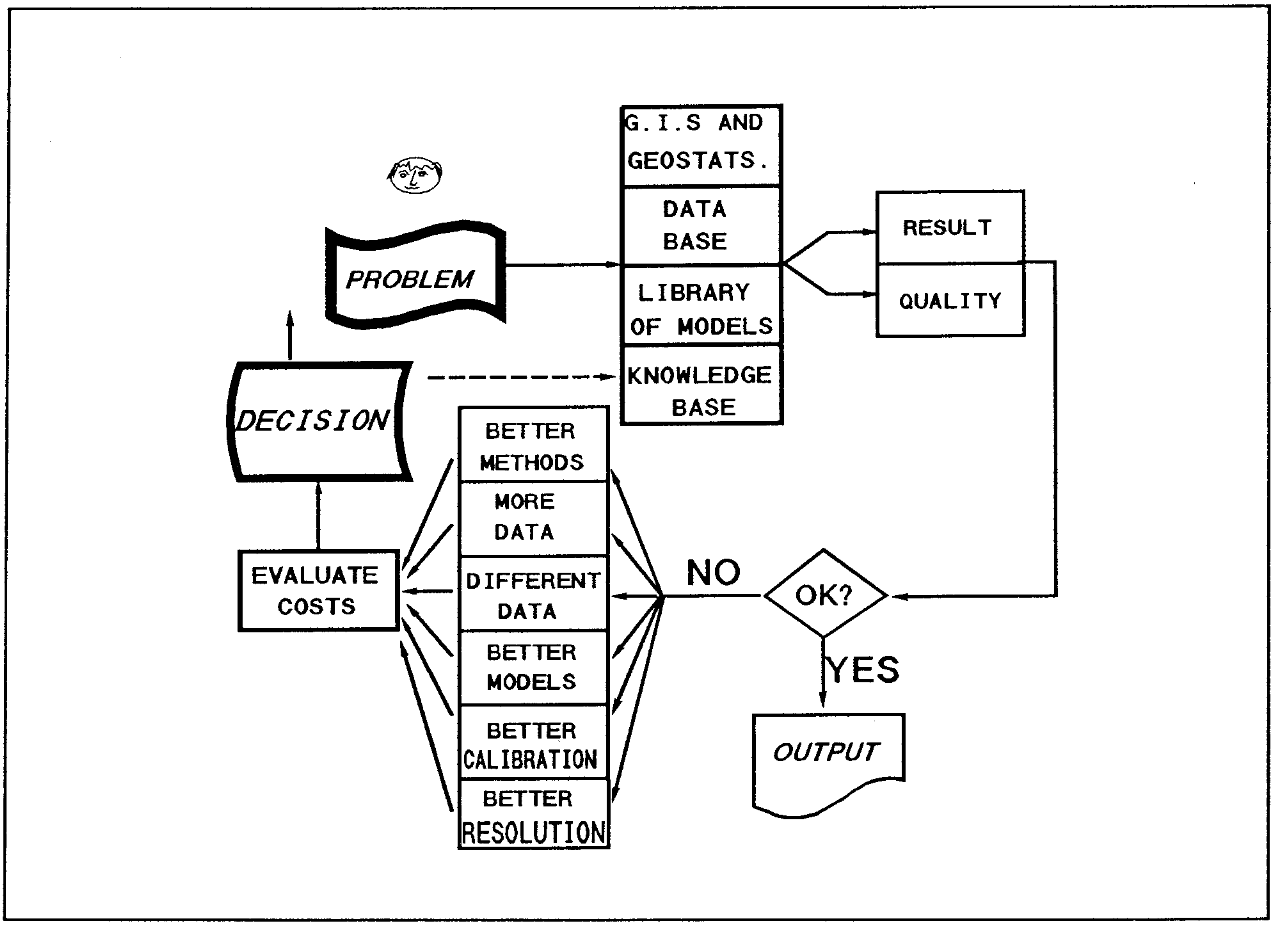


Figure 1. Flowchart for an intelligent GIS.

intelligence to assist a user to choose a suitable combination of GIS tools (data retrieval, statistical and geostatistical analysis), models and data to yield results of a required quality within a given budget; if this was not possible, then the GIS should suggest one or more strategies with which the user could achieve his or her goal. This idea is expressed in Figure 1.

Because GIS lack this intelligence the user often has little idea on how to achieve optimal results. Expertise about how to use GIS and how to use models of climate, groundwater movement or crop simulation needs to be formalized in such a way that it could be included as an important part of an intelligent GIS. A really intelligent GIS would be able to carry out error propagation studies *before* a major data crunching operation to estimate if the methods and data chosen were likely to yield the results intended. It would report to the user where the major sources of error come from and would present him with a set of options with which he could achieve better results. These options are:

a) use better methods (e.g. use co-kriging interpolation instead of simple kriging or just averaging, or use mathematical models instead of logic)
b) collect more data
c) improve the spatial and/or temporal resolution
d) collect different data
e) use better models
f) improve the model calibration

Each option would be accompanied by an estimate of the costs so that rational decisions could be taken. The system would also be able to indicate situations in which the results were *much better* than expected: in these cases important savings on data collection and processing could be made without serious loss of information.

Error estimation and error propagation.

Until recently there were few general tools for estimating error propagation in GIS. Of course, the stochastic theory of error propagation is well established (Taylor 1982) and forms a good starting point. Today, however, all the elements for studying error propagation are available. Modern statistical packages provide information on the standard errors of regression coefficients and residual standard errors and cheap powerful computing provides the means for Monte Carlo simulation studies at reasonable costs. Geostatistical interpolation methods provide estimates of kriging interpolation errors and databases are large enough and flexible enough to store quality control data along with the original data.

All these elements have been brought together in a single computer package by Heuvelink, Wesseling and co-workers at the University of Utrecht. The package is called ADAM (Wesseling and Heuvelink 1991) and it provides several methods for following the propagation of errors through modelling with GIS (Heuvelink 1993). ADAM can be "bolted on" to any raster or vector GIS because it uses the command language of the GIS itself for carrying out computations. At the University of Utrecht it is mainly used with a computational raster program (PC-RASTER) and associated geostatistical software (SPIL - Pebesma 1992), but has also been linked to ARC-INFO. A link to Intergraph's ERMA software is being discussed.

ADAM provides four main methods for analysing error propagation - first and second order Taylor series approximation, Rosenbleuth's technique, and Monte Carlo simulation. Applications to empirical transfer modelling, logical and fuzzy modelling, calculation of slope and aspect, and error propagation in runoff modelling with GIS are given in Heuvelink (1993), Heuvelink et al (1989), Heuvelink and Burrough (1993), de Roo et al (1992).

The presentation of quality information to a user.

This is the most difficult and uncertain part of the whole process. We await the outcome of the NCGIA competition on presenting information on data quality graphically.

ERROR PROPAGATION IN GIS - A PRACTICAL EXAMPLE.

Aim and context of the study.

The pollution of alluvial soils with sediments contaminated with heavy metals is a particular problem on the floodplains of many rivers in industrialised lands, and this is particularly so for the riverine areas of the River Maas in the southern part of the Netherlands. Mining and other activities in historical and recent times have caused the Maas to carry large amounts of heavy metal-laden silt, which is deposited on river banks and floodplains during flooding. The problem is not unique to the larger rivers but also occurs in smaller streams in the same area (e.g. Leenaers et al 1989). Because these floodplains are used for agriculture there is a real danger of livestock and crops being contaminated with lead, zinc, cadmium, copper and other undesirable materials. According to Dutch law, soil with amounts of heavy metals exceeding critical levels must be excluded from certain kinds of land use or, in extreme cases, must be sanitised. Consequently there are good economic reasons for needing to know which areas are polluted and if the degree of pollution exceeds critical thresholds.

Mapping chemical properties of the soil requires sampling, laboratory determination and a method for extrapolating from point measurements to areas. The most expensive items are labour costs for collecting samples in the field, and costs of laboratory analysis. Data storage and processing also cost money, but are becoming increasingly small components of the total cost. Therefore the main cost components determining whether a site is polluted are a) costs of sampling, and b) costs of analysis. The total cost depends directly on the number of samples taken and the costs incurred per sample.

The success of the mapping programme depends not only on the number and cost per sample, but also on the method used for mapping. Mapping options range from standard GIS polygon maps with fixed attributes (the choropleth map model), through interpolation using kriging or other techniques to empirical models based on an understanding of the flooding and deposition process (regression models). The last is potentially attractive because if a cheap-to-measure attribute is strongly correlated with an expensive-to-measure heavy metal concentration, then many observations of the cheap-to-measure attribute may serve to model the spatial distribution of the desired property at a suitable level of spatial resolution, but without incurring large costs. Methods of multivariate (co-)kriging use information from supporting attributes to improve the interpolation of expensive-to-map continuous attributes. Therefore, there are many ways in which the heavy metal concentrations can be mapped. The question is: which combination of method, number of data and kinds of data gives best value for money, and is that result good enough for the user?

Site, data, correlations and prediction and validation data sets.

The study area is situated west of the town of Stein in South-Limburg, the Netherlands. The area is about 5 km^2: it is almost totally inundated when the discharge of the Maas exceeds

1,500-2,000 m^3/s. Landuse is pasture and crops. Three soil types are present: 'ooivaaggronden' with and without lime and 'radebrikgronden'.

One hundred and fifty five topsoil samples were collected at sites located by stratified random sampling based using information on elevation relative to the river and geomorphology.
Bulk samples were collected by taking a small amount of soil at approximately 10 cm depth at 10 places located within a radius of 10 meters. This resulted in a sample of approximately 2-3 kg per site (Rikken and van Rijn, 1993).

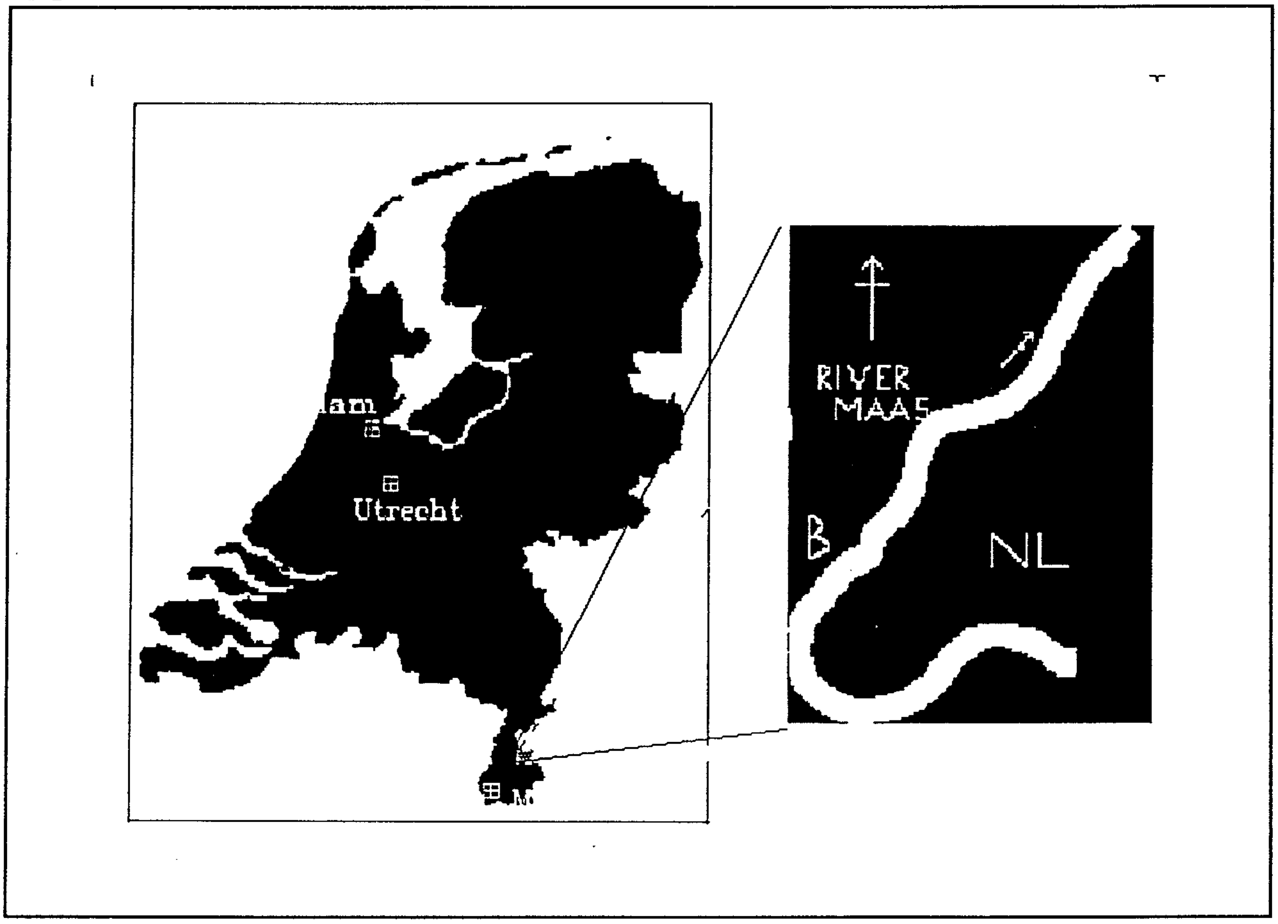


Figure 2. Location of the study area.

Previous research (Leenaers et al 1989) showed that heavy metal concentrations in floodplain soils in the Limburg area is related to relative elevation on the floodplain, and also to distance from the river bank. These properties are possible candidates for cheap surrogates for mapping heavy metal concentrations. This study uses the following data: easting and northing (accurate to c. 1m), zinc content (laboratory determination on sample in ppm), elevation with respect to the river (determined by GPS system in the field and corrected for river fall), and the distance in metres of the sample from the nearest point on the river bank.
Preliminary analysis showed that zinc content and distance to the river were strongly log-normally distributed, so these attributes were transformed to natural logarithms. All statistical analysis proceeded with the log-normally transformed zinc and distance data.

Correlation analysis carried out on all samples showed that there was indeed a strong negative correlations between ln(zinc) and ln(distance) [-0.768], between ln(zinc) and relative elevation [-0.624] and positive correlations between ln(distance) and relative elevation [0.481]. Before proceeding further, the total data set was split randomly into two parts; 102 samples were used for comparing mapping methods and the remaining 53 were kept apart for validation. Statistical comparisons of the two sub-data sets suggested that they were not significantly different samples of the study area.

The range of prediction methods and their estimates of errors

The following mapping methods were used: (error estimation in brackets)

Choropleth maps.
a) the three-class soil map,
b) the flooding frequency map
(analysis of variance)

Kriging interpolation
c) ordinary point kriging of Zinc levels
d) multivariate kriging of Zinc with distance to river, and
zinc with distance to river and relative elevation.
(variograms, covariograms and kriging errors)

Empirical models (regression)
e) simple linear regression of zinc on distance to river
f) multiple linear regression of zinc on distance to river and relative elevation
g) cartographic modelling to create distance zones from the river bank followed by simple linear regression (as e).
(all using regression coefficient errors, correlations, residual errors and error propagation using second order Taylor series approximation).

Figure 3 presents some examples of the maps and their associated errors.

Empirical models use the observed relation between zinc content and relative elevation and distance to the river. Methods e) and f) do this with respect to the measured values of elevation and distance at the sites used for soil sampling. Method g) calculates the effective distance of a site from the river from a digital raster map of the area. The SPREAD THRU option in cartographic modelling was used to create a non-linear buffer zone using the location of the river and a simple digital elevation map made by interpolating the elevation data. Although the DEM was made using GPS observations, in this area it could have been made more cheaply and just as effectively by digitising spot heights from a 1:10000 topographic map.

The multivariate point kriging also used the data from ln(distance) and relative elevation to improve the interpolation of ln(zinc). In the following analysis the costs and quality are evaluated for the situation in which the supporting data on distance to river and relative elevation were measured at all (102) sites.

Quality versus number of data points.
The prediction data set was divided into subsets of different size to investigate how prediction success varies with method and number of samples. These data sets included 102, 51, 27 and 14 samples of zinc determinations. When more than one subset of a given size was used (51, 27, 14 subsets) the results of the cost-quality studies were pooled.

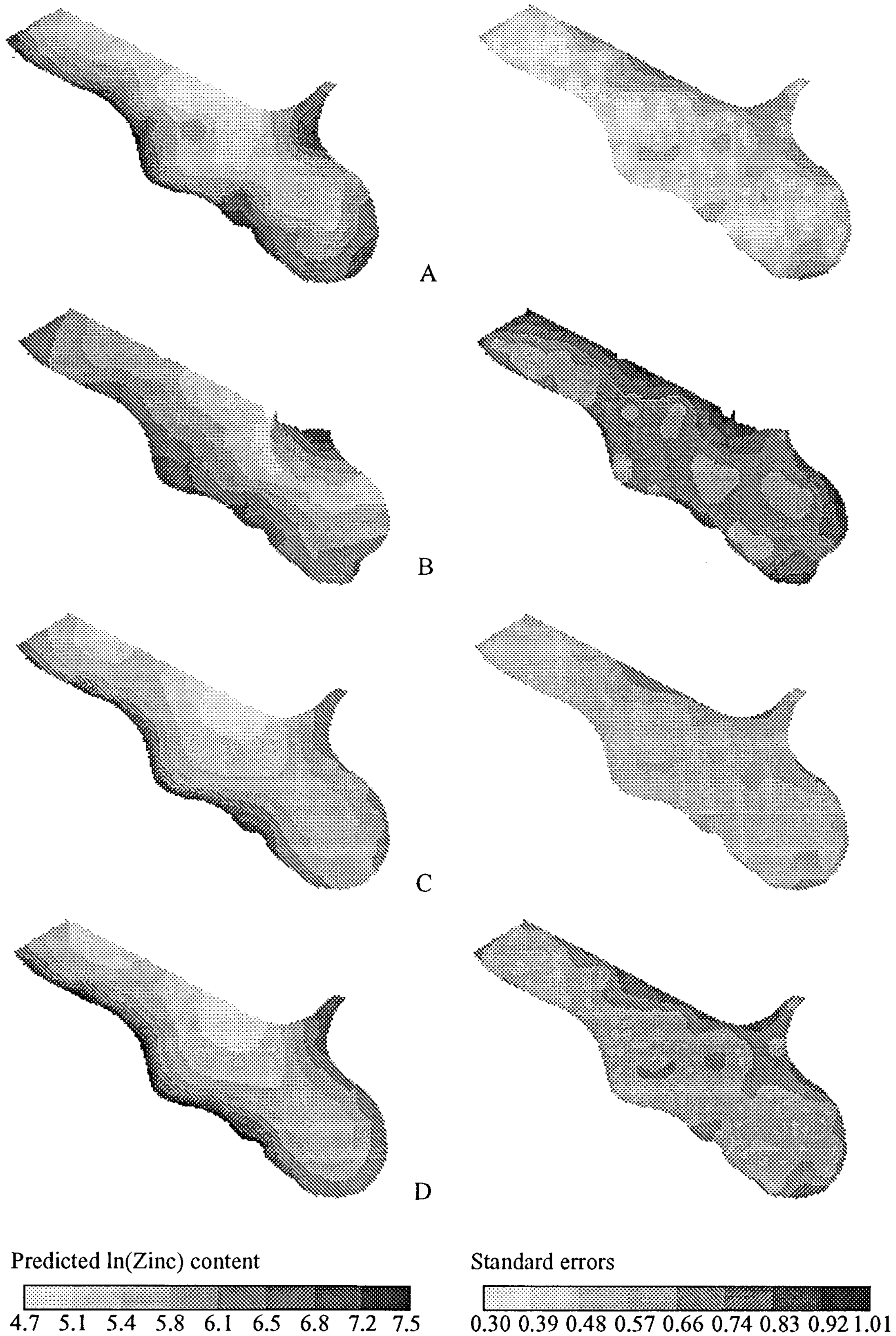


Figure 3. A) ln(Zinc) point kriging 102 sites; B) same, 27 sites; C) ln(Zinc) regression model 102 sites; d) same, 27 sites.

Costs of data collection.

Real costs of data collection and chemical analysis were sought from commercial environmental firms. Costs were estimated as follows:

sample taking in field - 52 units;
measuring altitude with GPS - 160 units
sample taking and altitude measuring together - 212 units
laboratory analysis for zinc - 100 units (150 for 4 elements)

(One cost unit is roughly equivalent to one Dutch guilder or AUS$0.7)

Costs of topographic maps for measuring distance to river were negligible. Computing costs for all methods were very low as methods are now very quick. Even multivariate kriging on a 486D 50 Mhz microcomputer took only 4 minutes to interpolate the 160 x 220 grid-cell map covering the area. Table 2 shows how these costs translate into campaign costs for sampling numbers plus methods (rounded to the nearest 500 units). Costs of data management and computing have not been included as these are roughly the same for all analyses.

Estimating the mapping errors.

The prediction standard errors (PSE) in the maps were estimated as follows:

a) Choropleth maps.
Data points were allocated to the class in which they were found and the mean and standard deviation of each class was computed. The best estimate for any point located in any grid cell in the map was then given by the unit average with standard error equivalent to the map unit standard deviation.

b) Kriging maps.
The standard error for point predictions was used, following normal practice. Optimal interpolation proceeded via variogram estimation and variogram model fitting, cross-validation and kriging using the SPIL program. Results were displayed with PC-RASTER.

c) Regression models and cartographic modelling.
Regression models, correlations and correlations between model coefficients and residual errors were computed using standard statistical methods. The ADAM package was used to create control files for error propagation using second order Taylor series approximations. Model results and error maps were computed and displayed using PC-RASTER

Criteria of quality: results.

Three criteria of quality were used. First, the mean PSE was calculated over all interpolated cells in the map. This gives a measure of how well the mapping method claims to be operating.

Second, each method was used to predict the value of ln(zinc) at each of the 53 data points in the validation set. The average prediction error (predicted - validation value) and the average root mean square error (SQRT((prediction - validation)^2)/53) were computed for each map (RMSE).

Finally, the difference between the PSE and the RMSE were evaluated by computing the index (PSE - RMSE)/PSE. Positive values of this index would show methods that appeared to predict better than they claimed, whereas negative values should show up methods that are less successful than they make out. PSE, RMSE and the success index were plotted against the costs of each mapping procedure (Figure 4,5,6).

Systematic bias.
All methods gave maps with very limited systematic bias when compared with the validation data.

Table 2. Costs of data acquisition for the different mapping methods.

Method	# samples zinc	# attributes (# samples supporting attributes)	Kind of attributes	cost (units x 10^3)
soil &	14	1	ln(zinc)	2
floodfreq.	27	1	ln(zinc)	4
	51	1	ln(zinc)	7.5
	102	1	ln(zinc)	15.5
ordinary point	27	1	ln(zinc)	4
kriging	51	1		8
	102	1		15.5
simple linear	14	2(102)	ln(zinc)	2
regression	27	2(102)	ln(distance)	4
and cartographic	51	2(102)		7.5
modelling	102	2(102)		15.5
multiple	14	3(102)	ln(zinc)	18
regression	27	3(102	ln(distance)	20
and multivariate	51	3(102)	elevation	24
kriging	102	3(102)		32

PSE - cost comparisons.
Figure 4 shows that the choropleth maps, though not expensive, generally perform least well, except when fewer than 35 data points are used, when they claim to be better than ordinary point kriging. This is because it is difficult to compute a good variogram with less that 50 data points, even with the strong pattern shown in this area. Simple linear regression models using distance to river are more effective than the choropleth maps, though the cartographic algebra method claims to be better than the point-based regression on distance. This is because the distance to each cell is determined deterministically by cartographic modelling and the error in distance is assumed to be limited to half a cell size (10m). The flattening off of the curves for regression, both with point observations of distance and computed distance shows the limitation of the regression model to further improvements with increasing numbers of samples. Once 50-60 samples have been used to calibrate the model, there is little further improvement.

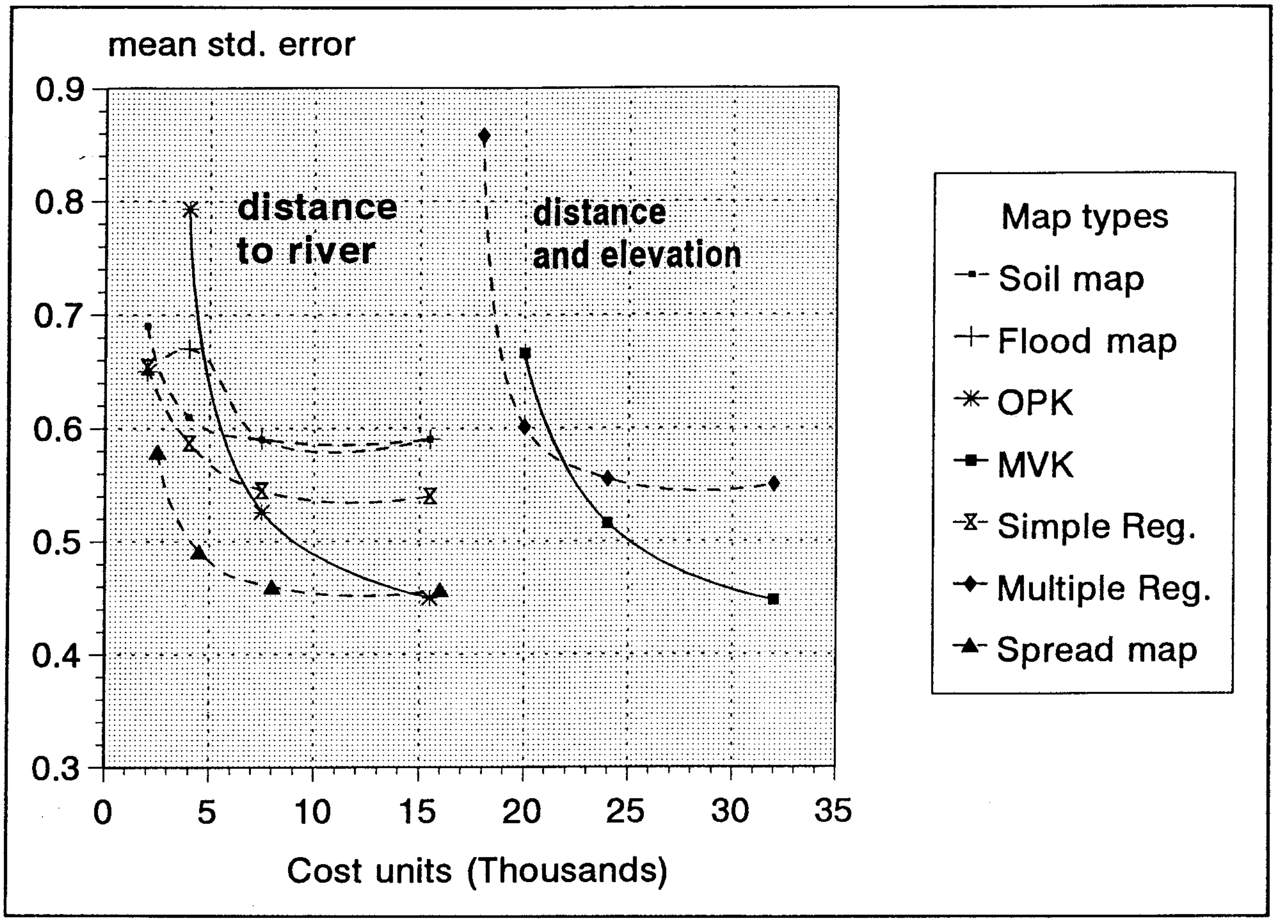


Figure 4 Cost-quality relations given by the prediction method.

The multiple regression model shows results similar to the simple regression model, but for more cost. This is because the relative elevation adds very little extra information in this case (the correlation with distance is large, and elevation measurements are very noisy yielding a variogram with a nugget variance that is c. 60% of the sill). Clearly the investment in elevation data by expensive GPS methods was not at all worthwhile in this area!

Both kriging methods continue to improve with increasing numbers of data, because of better variograms and a denser data net. Variograms cannot really be computed for less than 50 data points here, but for ordinary kriging a variogram was teased out of 27 data. Clearly such an interpolation is not worth much. However, once sufficient data become available, then ordinary kriging outperforms all other methods. Multiple kriging adds little extra information at greatly increased cost for the same reasons that multiple regression fails to yield improved estimates. In this case, the data exhibit the phenomenon of *intrinsic correlation* (Olea 1991) which means that the covariances and cross-covariances are proportional. This means that ln(Zinc) is *autokrigeable*, and the point kriging is the same as the co- (or multivariate) kriging. Kriging outperforms regression techniques once there are enough data because it is a local estimator and the regressions are global methods.

RMSE - cost comparisons

Figure 5 presents the results of the validation exercise which confirm the major features described above. Choropleth maps perform worst, with somewhat unexpectedly the flood frequency map being inferior to the soil map. The simple linear regressions on distance, both

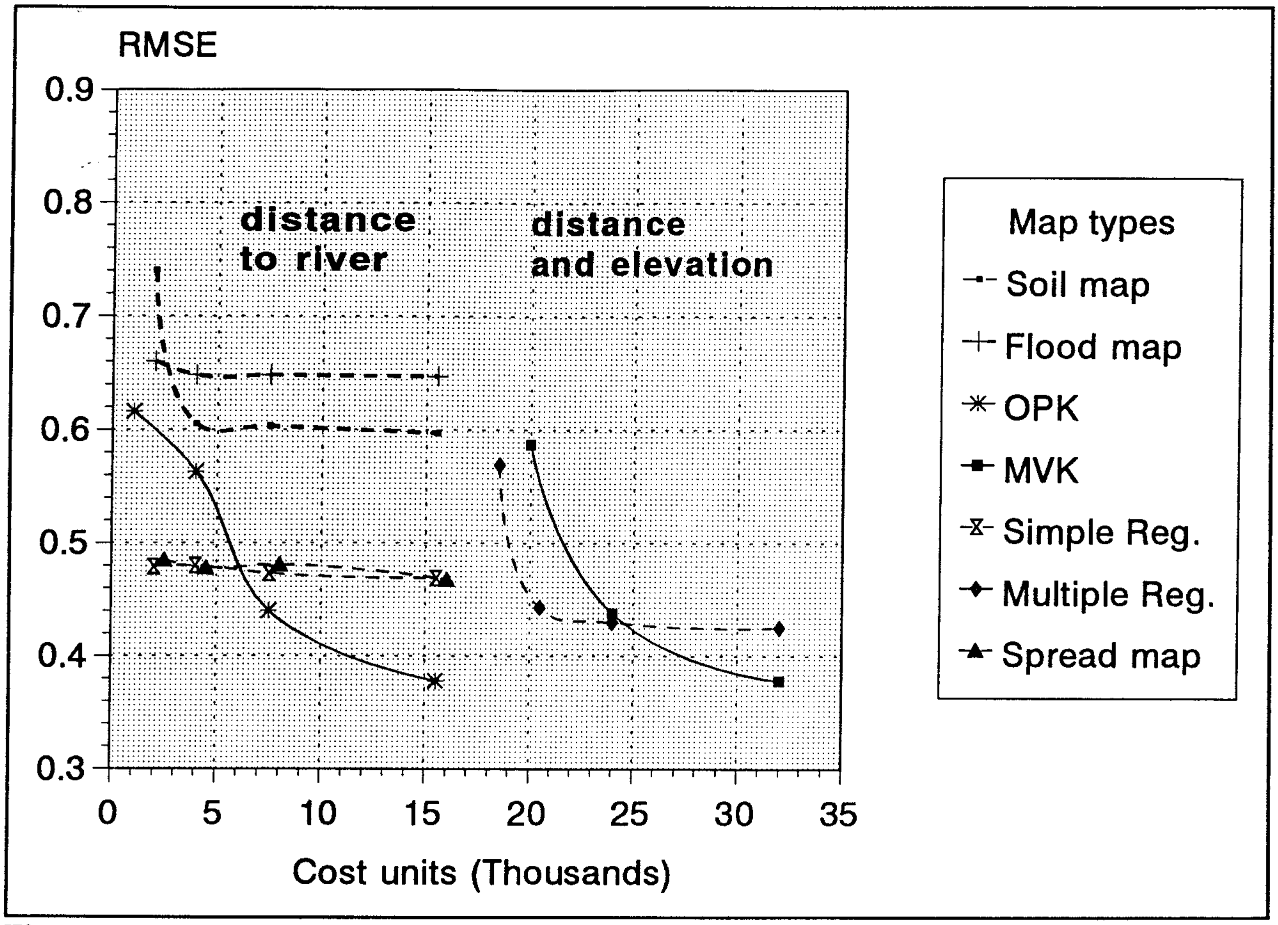


Figure 5. Cost-quality determined by 53 validation data.

interpolated from 102 data points and by cartographic modelling, return very similar results; multiple regression gives slightly better results, but at great costs. Kriging methods outperform all others once data exceed c. 40-50 samples, and multivariate kriging gives no extra advantage for the cost of the extra, redundant, data on elevation.

Prediction success index.

The prediction success index (Figure 6) shows that the choropleth maps perform relative to the validation set as expected once the number of samples exceed c. 30 for the area. The cartographic modelling regression method proves to be somewhat too optimistic, except when data numbers are small. In contrast, all regression and kriging methods appear to perform better than expected, though the comparison may be somewhat biassed because the validation data set may not have had many sites near the edge of the study area where interpolation errors are known to be large.

Discussion and conclusions.

The conclusions of this study suggest that all methods give reasonably unbiased predictions and believable maps. However, if one can only afford few measurements of an expensive attribute, then a sound empirical model will yield the best results (up to the limit of the residual error of the model). More complicated models may have lower residual errors (i.e. they describe the process better) at the cost of requiring more data. Cartographic modelling (SPREAD) may be useful for determining distance data geometrically from cheap digital elevation models.

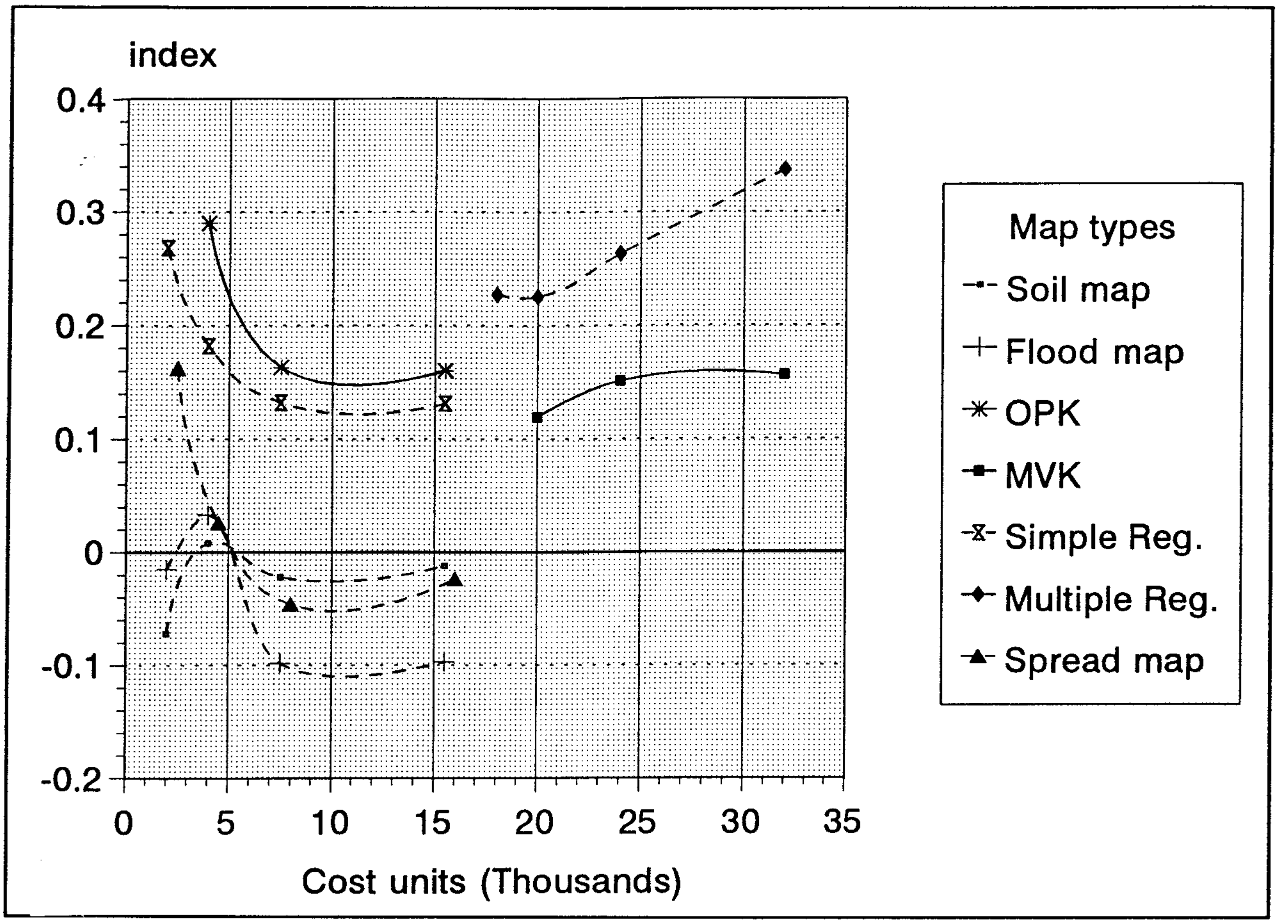


Figure 6 Cost-quality prediction success index.

Once the data are sufficient to support variogram calculation, kriging outperforms transfer models for spatial prediction. Multivariate kriging is of dubious benefit when extra data are correlated with others and when these surrogate data are expensive to measure. Cross-variogram fitting is a difficult and not entirely objective procedure that requires special expertise. Even with sufficient supporting data, multivariate kriging can only be carried out properly if there is a sufficient base for estimating the variograms and cross-variograms. If the kriging mapping is to be extended to similar areas then investing in sufficient samples to determine the variograms may be worthwhile, but needs to be demonstrated. Areas upstream or downstream of the present site would have different absolute levels of heavy metal pollution because of their different distances from the sources of pollution.

Error propagation methods permit the analysis of results to see which source of information (model or independent attributes) contribute most to the final error. If the model is not at fault, knowledge of the variogram can permit estimation of sampling densities necessary to attain a required level of error (McBratney and Webster 1981). Note that we still need to find out if any of the methods yield results of the calibre needed for managing the heavy metal problem of the sediments and this needs to be worked out in terms of tolerance levels for different kinds of land use.

The results suggest that the following procedures will lead to the best use of resources:
1. Use any sensible exploratory statistical procedure to see if there is a statistical relation that could lead to a simple empirical model. if there is, then this can provide a quick zoom-in on

the problem for limited cost. Don't spend money collecting correlated variables - this leads to expensive redundancy.
2. Use reconnaissance methods (field survey, nested analysis of variance) to get an idea of the spatial covariance structure of the desired attribute. Sample to establish this spatial dependence. If funds are limited, use spatial transfer models for continuous spatial processes (flooding, deposition, erosion); reserve non-locational (point) models for location-specific processes. If funds are sufficient, then sample systematically at the recommended level to map with kriging to yield a required standard error.

Note that this problem concerned a relatively simple pattern and process. Flooding is the main cause of a site being polluted and this is ultimately controlled by the landscape geometry. Short-range variation in heavy metal content was removed by bulking and homogenizing the soil samples. In many situations it may not be so clear cut. Raw data may have been obtained from small point samples and local spatial and temporal variability may be large, thereby obscuring long-range patterns. Such variation in physical and chemical properties is common in many young alluvial and periglacial soils or as a result of complex landscape interactions. For example, throughfall measurements indicate large variation in deposited chemicals over distances of metres at forest edges or under various kinds of trees (Draaijers 1993). Consequently there is a need to develop a general, formal understanding of links between models, data, costs and quality which can be used in intelligent GI-systems to advise users how to model the environment cost-effectively.

Acknowledgements.
Thanks for collecting the data to M. Rikken & R. van Rijn, and their supervisors H. Leenaers and A. de Roo. Thanks for providing the computing tools and for assistance with data processing and analysis to C. Wesseling, G. Heuvelink, E. Pebesma and V.Jetten.

REFERENCES.

Barrow, J.D., (1991). *Theories of Everything*. Vintage, London. 223pp.

Burrough, P.A., (1986). *Principles of Geographical Information Systems for Land Resources Assessment.* Oxford University press, Oxford. 196pp.

Burrough, P.A., (1992a). Are GIS data structures too simple minded? *Computers & Geosciences* 18: 395-400.

Burrough, P.A., (1992b). Development of intelligent geographical information systems. *Int. J. Geographical Information Systems* 6: 1-11.

de Roo, A.P.J., (1993). *Modelling surface runoff and soil erosion in catchments using Geographical Information Systems.* Netherlands Geographical Studies No. 157, University of Utrecht, 295pp.

Roo, A. de, Hazelhoff, L. & Heuvelink, G.B.M., (1992). Estimating the effects of spatial variability of infiltration on the output of a distributed runoff and soil erosion model using Monte Carlo methods. *Hydrological Processes* **6:** 127-143.

Draaijers, G.,(1993). *The variability of atmospheric deposition to forests.* Netherlands Geographical Studies No. 156, University of Utrecht, 208pp.

Heuvelink, G.B.M., (1993). *Error propagation in Quantitative Spatial Modelling: applications in Geographical Information Systems.* Netherlands Geographical Studies No. 15?, University of Utrecht, (in preparation).

Heuvelink,G.B.M. & Burrough, P.A., (1993). Error propagation in logical cartographic modelling using Boolean methods and continuous classification. *Int. J. Geographical Information Systems* 7: (in press).

Heuvelink, G.B.M, Burrough, P.A. & Stein, A., (1989). Propagation of errors in spatial modelling with GIS. *Int. J. Geographical Information Systems* 6: 1-11.

Leenaers, H, Burrough, P.A. & Okx, J.P., (1989). Efficient mapping of heavy metal pollution on floodplains by co-kriging from elevation data. In: *Three dimensional applications in Geographic Information Systems*, ed. J. Raper. Taylor and Francis Limited 37-50.

McBratney, A.B. & Webster, R., (1981). The design of optimal sampling schemes for local estimation & mapping of regionalized variables: 2 Program & examples. *Computers & Geosciences* **7**: 335-65.

Maguire, D.J., Goodchild, M.F. & Rhind, D.W. (1991). *Geographical Information Systems.* Longmans, Cambridge. 2 vols.

Olea, R.A., 1991. *Geostatistical Glossary and Multilingual Dictionary.* International Association for Mathematical Geology Studies in mathematical Geology No. 3, Oxford University Press, New York, 177pp.

Pebesma, E.,(1992). SPIL - a program for spatial interpolation. Department of Physical Geography, University of Utrecht/Dutch National Institute for Public Health & Environment.

Stewart, I., (1989). *Does God play dice? The Mathematics of Chaos.* Blackwell, Oxford. 317pp.

Taylor, J.R., (1982). *An introduction to error analysis.* OUP, Oxford. 270pp.

Rikken M.G.J. & van Rijn R.P.G., (1993). SOIL POLLUTION WITH HEAVY METALS. An Inquiry into Spatial Variation, Cost of Mapping and the Risk-evaluation of Copper, Cadmium, Lead and Zinc in the Floodplains of the Meuse West of Stein, The Netherlands. M.Sc. Dissertation, Department of Geography, University of Utrecht.

Wesseling, C.G. & Heuvelink, G.B.M., (1991). Semi-automatic evaluation of error propagation in GIS operations. *Proc. EGIS'91.* Ed. J.J. Harts, H.F. Ottens & H. Scholten. EGIS Foundation, Utrecht, 1228-1237.

Whitehead, A.N., (1953). *Science and the Modern World.* Cambridge University Press.

Line Intersect Sampling for Map Accuracy Assessment

M.S. Causley
BHP Engineering
Land Technologies Division
26 Atchinson Street
Wollongong NSW 2500

and

A.K. Skidmore
Centre for Remote Sensing and GIS
School of Geography
University of NSW
PO Box 1, Kensington NSW 2033

Abstract

The issue of assessing map accuracy is one which has changed focus with the development of geographical information systems (GIS). The ease in which geographical data may be manipulated and analysed within a GIS, increases the need to understand the error associated with mapped data. Furthermore, there is a requirement to develop economical and relevant methods assessing map accuracy. This report implements a new assessment technique, line intersect sampling, proposed by Skidmore and Turner (1992) which incorporates an error band model to described the error associated with map boundaries. In addition, a sensitivity analysis of the method is conducted, examining the variation in map accuracy estimate with a change in an epsilon error band distance. A comparison is made between map accuracy estimates determined using line intersect sampling and more conventional point sampling methods for a forest type mapping application in north-eastern New South Wales. The results highlight conservative estimates obtained using Line Intersect Sampling, which are largely attributed to the stringent testing criteria used. Whilst the map accuracy estimates are considerably lower than conventionally determined estimates, it is argued that the line intersect sampling technique provides a more indicative assessment of map accuracy.

Introduction

While the development of geographical information systems (hereafter termed GIS) is progressing at an impressive rate, there are problems with this technology. One of these problems is the consideration of error associated with GIS data.

The issue of identifying map error has been addressed by several researchers (Chrisman, 1984, 1987, 1989; Goodchild, 1987, 1988, 1990, 1992; Veregin, 1989). Overall, there has been a general consensus reached as to the definition of the prime components of error in survey cadastre or human mapping. The components being locational and attribute error. In positional terms, accuracy may be estimated by a Root Mean Square Error (RMSE). The RMSE is a cumulative value defined by the square root of the average of the squared discrepancies (ASPRS, 1988). The discrepancies are the differences between 'mapped' and 'true' co-ordinate location. However, such estimates do not give any indication as to the quality of the map classification.

Methods for assessing map classification accuracy usually follow conventional point sampling/confusion matrix procedures. For this method to be successful, accurate and economical field sample data are required to analyse the integrity of the map. Skidmore and Turner (1992) proposed a technique for assessing the accuracy of GIS maps utilising line intersect sampling and epsilon error band theory. Line intersect sampling is a technique which collects land cover boundary information along randomly located traverses in the field. The field data are then compared to the boundaries plotted on the image, in this case, a vector map. The epsilon error band theory, initially proposed by Perkal (1966), describes the uncertainty associated with the cartographic representation of class boundaries. Perkal (1966) defined a distance (epsilon) about a cartographic line as a means of objectively generalising the line. The theory was further adapted by Blakemore (1984), who used the epsilon distance to describe the width of an error band about a digitized line of a polygon boundary. Skidmore and Turner (1992) modify the error band theory adopted by Blakemore (1984). The aim of the technique proposed by Skidmore and Turner (1992) was to provide simple but effective field measurements to obtain an estimate of map accuracy. The accuracy assessment was based on a "boundary error" ratio, which is determined by the ratio of correctly located class boundaries with respect to the total boundary length.

The broad aim of this study was to perform a sensitivity analysis of the change in map accuracy, using line intersect sampling, with a change in the epsilon distance. As a consequence of this analysis, the magnitude of the epsilon band distance at which no further increase or decrease in the accuracy is discernible, is of particular interest. Also of interest is the difference in the map accuracy estimated by two different definitions of boundary 'correctness'.

Method

Line Intersect Sampling

This study implements the line intersect technique to determine the accuracy of a forest type classification map of the Chaelundi State Forest near Coffs Harbour, in northern New South Wales. The vector GIS base map (1:15000) was produced by the Forestry Commission of NSW from 1:15840 aerial photography. Aerial photograph interpreters classified the forested areas in accordance with guideline definitions outlined by Baur (1965). The forest type boundaries were captured by the Micromap digitizing package and then transferred to the ARC/INFO[1] GIS. Within the GIS, the linework was cleaned and the topology built.

In order to assess map accuracy, the line intersect method of sampling was used to determine differences between "true" and "mapped" location of the forest type boundaries. As discussed by Skidmore and Turner (1992), line intersect sampling has been

[1] ARC/INFO is a product of ESRI, Redlands California, USA.

used in forestry. Furthermore, it was shown that an estimate of map accuracy can be achieved by comparing boundary lengths between the mapped polygons and the true ground boundaries using the line intersect sampling technique. The ratio, known as the Boundary Error Ratio, is given by:

$$\text{Boundary Error Ratio} = \frac{\text{length of correctly located boundaries } (X_r)}{\text{total boundary length } (X_m)}$$

The total boundary length value is determined by in-field measurements. Transect line data is randomly generated, using initial ground co-ordinate locations, bearing and distance. Once the points have been located in the field, the transect lines are traversed, noting the changes in the forest cover classes. A count is made of the total number of class changes (*'m'*) that occurred. The total boundary length and variance are determined using formulae derived by De Vries (1986). It has been shown that the total boundary length of polygons per unit area (boundary density) of a map is equivalent to:

$$X = \frac{\Pi . m}{2 . L}$$

where X = estimated boundary length of polygons per unit area,
m = the number of intersections of the transect lines with the polygon boundaries, and
L = the total length of the transect line(s).

The variance of the boundary length of polygons per unit area is given by:

$$\text{var } X = (\Pi / 2 . L)^2 . m$$

By comparing the field data with the mapped data, a count can be made of the number of correctly located boundaries (notated as *'r'*). Correctly located map boundaries are defined as those in which a generated error band (epsilon distance) about the 'mapped' boundary line coincides with the true location, and the attributes on either side are correctly labelled. The magnitude of the epsilon distance error band is representative of the accumulation of factors which contribute to the presence of error in the map, including errors created during map compilation. From the counts of 'true' boundary intersections (*m*) and coincident 'mapped' boundaries (*r*), an estimate of map accuracy is determined by calculating the total 'true' boundary density (X_m) and the 'mapped' boundary density (X_r). The map accuracy estimate is then represented by the ratio of X_r to X_m as follows:

$$Xt = (X_r / X_m) . 100$$

Fieldwork

A number of transect starting co-ordinates and compass bearing directions were randomly generated on a 1:15000 base map within the case study area, and transferred to aerial photographs. Transects lines were traversed through the forest using a hand-held compass and measuring tape to determine the location of ecotonal changes in the forest types. The chainage position of the forest changes were recorded together with the forest types (Baur, 1965). Some difficulties arose at transitions between forest types where, sample plots were used to determine the dominant species.

The accuracy standard of the reference data from which the assessment is determined should be of a higher order of accuracy compared with the mapped data. As the forest type mapping was compiled at a scale of 1:15000, the field work would be required to be accurate to within ±7.5 metres (±0.5 mm at mapping scale) to meet lower spatial map accuracy standards (ASPRS, 1988). The hand held compass and measuring tape techniques used to collect the ground data would be expected to easily achieve such standards.

The minimum of number samples needed to achieve a 95 per cent confidence level of the accuracy assessment was determined by assessing the trend of the forest types along the first transect. From the number of species changes in comparison to the length of the traversed transect it was calculated that nine transects of 900-1000 metres in length would be required to obtain an adequate sample. The first transect traverse length was L = 930 metres, along which nine forest type changes were delineated (m = 9). The boundary length of polygons per unit area and the associated variance were estimated as:

$$X = 0.015\,201$$

$$\text{var } X = 0.000\,025$$

The number of transect line samples required to meet the 95 per cent confidence level is determined such that:

$$n = \left(\frac{t^{0.975}_{m-1} \cdot CV(X)}{E} \right)^2$$

Where,

	n	=	number of samples required;
	t	=	Student t distribution;
	CV(X)	=	coefficient of variance; and
	E	=	prespecified relative allowable error fraction.

(De Vries, 1986)

Therefore, the required number of transects was determined to be 9.

The prespecified relative allowable error fraction (E), provides a 95 per cent confidence interval of the accuracy for the determined value of X (Walpole and Myers, 1978). However, the value X will not be exactly equal to the true boundary length of polygons per unit area for the map. The difference between the estimated and true value of X is described by E (Walpole and Myers, 1978). De Vries (1986) discusses the criteria used to determine suitable magnitudes of E for given applications. It is stated that a small E value implies a narrow difference margin between the estimated and true X value. Unfortunately, to achieve such a precision is costly, that is, requiring a larger sample. Furthermore, De Vries (1986) explains that there is a need to balance precision requirements with the goals of the sampling assessment. It is argued that for applications where there is considerable reliance on an estimate being a close reflection of the true value, the allowable error E would be in the order of 0.05 (5 per cent). For more general planning estimates, such as for this study, a value between 0.1-0.15 (10-15 per cent) is applicable. Unfortunately, the data for three transects was not available and not able to be used for the accuracy assessment. As a consequence, there was a reduction in the stated confidence level (See Appendix).

The analysis of the mapped data using the line intersect sampling was implemented within the GIS environment. A direct comparison was made between the field data and the vector map using overlapping coverage layers. The field transect measurements were stored as separate data layers within the GIS. Using the coordinate geometry (COGO) functions of the GIS, a transect data coverage was created by initially generating a point coverage using the randomly selected co-ordinate points. From these co-ordinate locations the field measurements were input into the coverage using an open traverse routine. By geo-referencing the field data using the same Australian Map Grid (AMG) co-ordinate system to that of the map data, correlation was achieved when the data layers were superimposed. To make the comparison of the two data layers easier, forest type classes from the ground sampling were attached to the transect coverage.

The epsilon distance error band required for the accuracy assessment was generated by creating buffer zones about the linework of the map data layer. A buffer zone layer was created for each separate epsilon distance used. The epsilon distances commenced at a distance of 5 metres and progressed to a maximum of 70 metres at 10 metre intervals.

Line Intersect Sampling: A Sensitivity Analysis

The vector map data and the transect field data were compared using sensitivity analysis to examine the variation in the accuracy estimate with a change in epsilon distance. In addition, the analysis of the accuracy estimate was produced in reference to two criteria categories. Two measures of accuracy were established to define different states of boundary correctness viz.

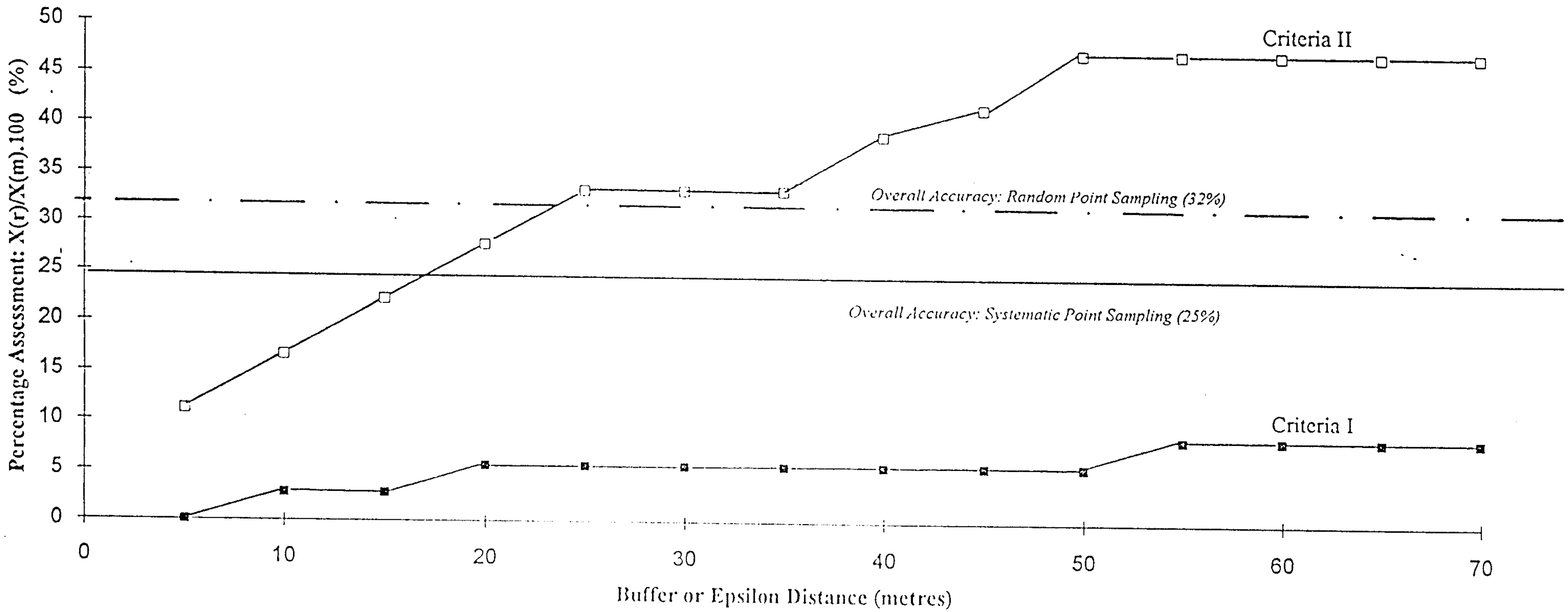
(i) *Criteria I* -

A boundary was considered 'correct' if:

(a) the field measured land cover change were located within the generated epsilon distance error band about the mapped boundary; and

(b) the attributes either side of the mapped boundary are correctly labelled.

Figure 1: Accurcy Assessment - Line Intersect Sampling Sensitivity Analysis and Point Sampling Results



(ii) *Criteria II* -

A boundary was considered 'correct' if:

(a) the field measured land cover change were located within the generated epsilon distance error band about the mapped boundary; and

(b) at least one attribute either side of the mapped boundary is correctly labelled.

Criteria I imposes a very rigourous definition of boundary 'correctness' which leaves a narrow margin for error in the mapping (at a particular epsilon distance). As a result, the implementation of line intersect sampling using *Criteria I* yields a conservative very estimate of map accuracy.

The guidelines for boundary 'correctness' established in *Criteria II* are based on the concept that, in positional terms, the map may be considered to be correct. *Criteria II*, in some respects, attempts to differentiate between locational and attribute error.

Results

Line Intersect Sampling Sensitivity Analysis

A summary of results for the Line Intersect Sampling method is shown in Figure 1. The graph illustrates the two criterion used to determine estimates of map accuracy. In addition, Figure 1 illustrates accuracy estimates obtained using more conventional point sampling techniques. Figure 1 highlights the considerably lower estimates of line intersect sampling compared with point sampling. The following results are discussed in reference to Figure 1 which were calculated to a confidence level of approximately 90 per cent.

Line Intersect Sampling Sensitivity Analysis using *Criteria I*

Firstly, accuracy estimates were determine from line intersect sampling technique at the various epsilon distances using the *Criteria I* boundary 'correctness' definition. This criteria enforces a strict requirement that both locational and attribute components are correct. Figure 1 highlights the influence of the stringent criteria used. At no stage did the map accuracy rise above 10 per cent (an extremely low value). Although, only low estimates of map accuracy were obtained, a distinct pattern of three levels of accuracy can be seen with, the first estimate of map accuracy achieved at a buffer distance of 10 metres. The next estimate level was determined at a buffer distance of 20 metres. The 5.55 per cent estimate remained until a buffer distance of 55 metres was used. It was at this stage that the maximum estimate reached 8.4 per cent.

Line Intersect Sampling Sensitivity Analysis using *Criteria II*

The second sensitivity analysis was performed using the *Criteria II* boundary 'correctness'. As shown in Figure 1, the percentage accuracy of the map ranged from 11 per cent to a maximum of 47 per cent. The results are characterised by two distinct plateaus. Initially, the percentage accuracy increased linearly, from a buffer distance of 5 metres to a distance of 25 metres. At this first plateau, the estimate levelled out at 33 per cent. The accuracy estimate again increased in an almost linear manner from a buffer distance of 35 metres to a buffer distance of 50 metres. It is at this second plateau that the maximum accuracy assessment was reached.

It should be noted that when the epsilon distance was greater than 40 metres, there was a reasonable degree of difficulty in interpreting the data. At these distances many of the smaller polygons were engulfed by the zones of uncertainty of the larger polygon areas. Depending on the type of information that is classified, the point at which there is no further increase in estimated accuracy will vary. For classified maps such as forest type coverages this value will be reasonably large due to the size and shape of the classified polygons.

Discussion of Line Intersect Sampling Results: An Overview

The low accuracy determined in this case, is largely attributed to error in the mapping itself. The *Criteria I* definition of accuracy in some respects, may be regarded as an unreasonable interpretation of map accuracy. Chrisman (1989) discussed an important consideration of map accuracy assessment being the decision about which component of a mapping coverage should take logical precedence: the spatial description or the attribute. Chrisman (1989) maintained that for resource mapping such as soil or vegetation taxonomy, the map is a function of each polygon being assigned to one class or other. Therefore, he argued that for such applications, issues of positional accuracy may be considered prominent. Thus, an option used was here define *Criteria II* requiring both adjacent polygon attributes to be correctly labelled. While this criteria does not imply total accuracy, it may be considered to give a better indication of locational accuracy.

It may be argued that a comparison of results obtained using the two criteria is indicative of attribute error. From the results, the greatest difference in accuracy estimates between the two criterion is over 40 per cent; a significant amount that may be largely accredited to attribute errors. This assumption is made due to the fact that *Criteria I* stipulates absolute locational and attribute correctness, while *Criteria II* only enforces locational correctness to the fullest. From comparison of the field data with the mapped data, it was apparent on some transect lines that the accuracy estimate could have been dramatically improved if some of the mapped forest type classes were directly swapped, suggesting high attribute error.

One explanation of the low estimate of map accuracy obtained using the line intersect sampling technique may be that the field measurements delineated many more forest type changes in comparison to the photo interpreter. Naturally, there is a requirement to match the parameters of the testing method with the characteristics of the data. For example, the precision of the field measurements and estimation of an epsilon distance should be determined by considering factors such as the mapping scale, the type of information being mapped, map production methods and the map's application.

Sensitivity Analysis of Line Intersect Sampling and Epsilon Distance

The results of the sensitivity analysis are characterised by a distinctive 'stepped' pattern for both assessment criterion. The stepped pattern is linked to increases in epsilon distance as the estimate of accuracy can only remain constant or increase with increasing epsilon distance. Once a boundary has been determined to be correct, an increase in the epsilon distance will have no further influence on its accuracy. Table 1 summarises the buffer distances at which the distinctive plateaus occurred on Figure 1. From examination of the 1:15000 vector map, the minimum polygon width represented was in the order of 2 millimetres, which corresponds to a ground distance of 30 metres. The results show a steady increase in the accuracy estimate prior to the first plateau. The first plateau occurs at an epsilon distance approximately equivalent to the minimum polygon width represented on the vector map.

Table 1: Occurrence of Plateaus in Results (*See Figure 1*)

	Plateau 1	**Plateau 2**
Buffer/Epsilon Distance (metres)	20-25	50-55
Map Distance Range (mm)	1.3-1.7	3.3-3.7
Average Map Distance (mm)	1.5	3.5

From examination of the mapped data and the results, it is not envisaged that a third plateau would appear for larger epsilon distances. Regardless of an increased epsilon distance it was obvious that there would be no further intersections between the mapped and field data.

When considering the cartographic representation of the polygon boundaries, it is important to realise that the pen width used to draw the boundary lines is often in the order of half a millimetre. At a mapping scale of 1:15000 this represents 7.5 metres on the ground. Accordingly, the very low estimates achieved for the smaller band widths are not unexpected. A minor deviation of the pen from the 'true' boundary has implications of significant error estimates, but this error would reduce at larger epsilon distances. However, examination of Figure 1 highlights that for *Criteria I* the error could not be attributed to the pen width as error stays high. Even at the largest error band distances there is little deviation from the low accuracy estimate.

Investigation of the methodology and theory of the line intersect sampling technique, suggests that there will be a known maximum estimate achievable for a particular assessment application. This maximum estimate will be dependent on the number of land cover changes found in the field and the number mapped along the transect lines. Table 2 outlines the number of intersections between polygon boundaries and transect lines obtained by field measurement and recorded on the map.

Table 2: Summary of Boundary - Transect Intersections

Transect Number	Number of Intersections between Boundary & Transect Line		Number "*m*" and "*r*" coincident boundaries from Line Intersect Sampling			
	Mapped	Field	*Criteria I*		*Criteria II*	
			Lower Buffer (5 metres)	Higher Buffer (70 metres)	Lower Buffer (5 metres)	Higher Buffer (70 metres)
4	4	6	0	0	0	1
5	6	5	0	1	0	1
16	5	5	0	0	1	4
19	5	4	0	1	0	1
20	5	7	0	0	2	3
25	7	9	0	1	1	7
Total	**32**	**36**	**0**	**3**	**4**	**17**

Because the technique bases its estimate on a total polygon length per unit area using the number of intersections along the transects for both the mapped and "true" field data, the maximum achievable estimate of accuracy will be equivalent to:

$$\text{Potential Maximum Accuracy} = \frac{\text{Boundary length / unit area for map}}{\text{Boundary length / unit area for field}} = \frac{X_{\text{map (total)}}}{X_{\text{field (total)}}} = 89\ \%$$

Therefore, if the mapped information matches the field measured data (albeit fewer polygons on the map), the highest percentage obtainable using line intersect sampling would be approximately 89 per cent. It is noticeable for *Criteria I* that there is no correlation for three of the transect lines, regardless of the buffer distance (epsilon distance) used. In a similar vein, there are situations whereby, for a particular transect, the potential accuracy is automatically reduced as there were fewer class changes found in the field compared to the mapped data. For example, Transects 5 and 19.

Another feature highlighted by Table 2 is the influence of the respective criterion used to define "correctness". As stated, three transects using *Criteria I* did not register one intersection between the mapped and field data. Two of these transect (Transects 16 and 20) show a significant improvement in accuracy using *Criteria II*. This suggests some degree of classification or attribute error. The results obtained for Transect 4 indicate that there is a substantial difference between what has been mapped and what was located in the field.

Conclusion

An application of line intersect sampling proposed by Skidmore and Turner (1992) has been conducted to investigate estimates of map accuracy. The results were compared with more conventional point sampling methods, with line intersect sampling providing more conservative estimates. The line intersect method provides a more economical field measuring design and testing algorithm that is simple to implement. Furthermore, the utilisation of the error band model theory, enables the technique to account for errors associated with map production processes. The error band model also provides the analyst with the capability of designating an error distance (epsilon) about a line that is suitable to the scale, resolution and purpose of the mapping.

There has been some debate as to the viability of an epsilon error band to model the uncertainty associated with the representation of geographical data. It is suggested that GIS technology entices users into a false sense of database accuracy. As a result, there is a need to understand the uncertainty of digital databases. The use of an epsilon error band is justified for such applications, as it attempts to accommodate and model errors resulting from production processes.

The results obtained from the sensitivity analysis indicate that there is an apparent correlation between the determined accuracy estimates and the characteristics of the mapped information. The structural nature of the forest class polygons, being commonly long and narrow, result in an increased sensitivity of this type of mapping to the accuracy assessment technique proposed by Skidmore and Turner (1992). This fact is highlighted by the accuracy assessment estimates achieved using *Criteria I*. An additional assessment was conducted using *Criteria II* which removed the requirement of total attribute correctness. The difference of 30-40 per cent between *Criteria I* and *II* emphasized the poor classification quality of the mapping.

Examination of the line intersect sampling method and the results achieved using the method, also indicated the importance of field measurement techniques that suit the scale and resolution of the mapping. It was shown that the manner in which the field work is conducted may dramatically influence the overall accuracy assessment. If the fieldwork delineates may more forest class changes than was mapped, the likelihood of a lower accuracy estimate is increased.

For future research investigation, there is scope to use the method on different mapping applications and at different scales. From such investigations the viability of the method for particular applications and the sensitivity of the epsilon distance model to a change in mapping scale may be evaluated. Furthermore, there may be cause to investigate the influence of a variety of field measuring accuracies on the overall accuracy estimate.

Appendix

Confidence Interval Calculation

$$(n / t_{0.05,5}) = ((CV(X) / E)^2$$

where, for this study $n = 6$

$E = 0.1$

$t_{0.05,5} = 2.015$

Therefore,

$$CV(X) = 0.173$$

$$CV(X) = 1 / \sqrt{m}$$

Therefore, $m \approx 34$ intersections which closely approximates the 36 intersections determined in this case.

References

ASPRS, (1988) ASPRS Interim accuracy standards for large-scale maps, *Photogrammetric Engineering and Remote Sensing*, Vol. 54 No.7, pp.1079-1081.

Baur, G.N., (1965) *Forest types in New South Wales*, Forestry Commission of N.S.W., Research Note No.17, Forestry Commission of N.S.W., Sydney.

Blakemore, M., (1984) Generalisation and error in Spatial databases, *Cartographica*, Vol.21, No.2&3, pp131-9.

Chrisman, N.R., (1983) The Role of Quality Information in the long-term functioning of a geographic information system., *Cartographica*, Vol. 21 No. 2&3, pp.79-87

Chrisman, N.R., (1987) The Accuracy of Map Overlays: A Reassessment, *Landscape and Urban Planning*, Vol. 14 pp. 427-439.

Chrisman, N.R., (1989) Modelling error in overlaid categorical maps., In Goodchild, M., and Gopal, S., (eds) *The Accuracy of Spatial Databases*, Chapter 2, Taylor & Francis, London.

De Vries, P.G., (1986) *Sampling Theory for Forest Inventory*, Springer-Verlag, Berlin.

Goodchild, M.F., and Dubuc, O., (1987) A model of error for choropleth maps with applications to GIS, *Proceedings AutoCarto 8*, ASPRS/ACSM, Falls Church, Virginia, pp.165-74.

Goodchild, M.F., (1988) The Issue of Accuracy in Global Databases., In Mounsey, H., and Tomlinson, R., (eds) *Building Databases for Global Science*, Taylor & Francis, New York, pp.31-48

Goodchild, M.F., (1990) Modelling Error in Spatial Databases, *GIS/LIS '90 Proceedings*, Vol.1, pp.154-162.

Goodchild, M.F., Guoqing, S., and Shiren, Y., (1992) Development and Test of an Error Model for Categorical Data, International Journal of Geographical Information Systems, Vol.6 No.2, Taylor & Francis, London, Washington, pp.87-103.

Perkal, J., (1966) On the length of empirical curves., Discussion paper 10, Michigan Inter-University Community of Mathematical Geographer., Cited by Blakemore, M., (1984) Generalisation and error in Spatial databases, *Cartographica*, Vol.21, No.2&3, pp131-9.

Skidmore, A.K., and Turner, B.J., (1992) Map Accuracy Assessment using Line Transect Intersect Sampling, *Photogrammetric Engineering and Remote Sensing*, Vol. 58, No.10, pp. 1453-1457.

Walpole, R.E., and Myers, R.H., (1978) *Probability and Statistics for Engineers and Scientist*, 2nd Edition, MacMillan Publishing Co. Inc., New York.

DATA QUALITY ASSURANCE AND CONTROL: A PRODUCER POINT OF VIEW

Gustavo Dominguez
Australian Surveying and Land Information Group
Topographic Information Unit
PO Box 608, Dandenong 3175
Australia

ACKNOWLEDGMENTS

I wish to acknowledge that the concepts regarding the quality control process and the figure 2 of this paper are based on a presentation given by Geoff Lawford, head of the Validation and Testing Section, Australian Surveying and Land Information Group, Canberra.

ABSTRACT

In many different ways, data quality is an issue for every GIS user. This paper describes how the Australian Surveying and Land Information Group (AUSLIG) is implementing quality assurance in the GEODATA TOPO-250K production lines and the post production quality control process.

Considering quality as "fitness for purpose" or in other words "compliance with a set of requirements", one of the major issues for a data producer is to communicate to the client effectively all relevant information concerning the quality of data. The way in which this is achieved complying with the oncoming Spatial Data Transfer Standard (SDTS) is also discussed, as well as the inclusion of feature specific quality attributes.

INTRODUCTION

As technology improves and users become more sophisticated in their applications the data producer faces the challenge of changing old habits and procedures to build databases that meet the client's demands.

AUSLIG's customer focus approach involves the adoption of Total Quality Management (TQM) concepts. TQM is in practical terms achieved by the implementation of a Quality Assurance system (QA).

AUSLIG is seeking world wide third party certification of its quality system according to the Australian Standards AS3901-1987 "Quality Systems for Design/Development, Production, Installation and Servicing" and AS3902-1987 "Quality Systems for Production and Installation", through the internationally recognised organization Det Norske Veritas. This will position AUSLIG among the leading organizations in the land information industry.

The experience of other industries shows that the adoption of TQM improves productivity and sets the foundations for product excellence.

THE QUALITY OF DATA

In the GIS industry, there has been an ongoing debate about data quality over the years.

The concept of quality has usually been associated with the positional accuracy of the data, mostly because this property, unlike many others, can be easily measured by comparison with a higher accuracy source.

But the positional accuracy is not really the basis of the concept of quality. As figure 1 shows, an engineer working on a construction project will require the data to have an accuracy of 1m; he will use the road centre lines and perhaps will not look at the extension of intertidal flats. On the other hand, a land administrator will be satisfied with an accuracy of 200m; he will refer to place names but will not be interested in the minor roads within a built-up area.

In fact, the application requirements are the parameters against which the user measures the goodness of data. Therefore, the true meaning of quality is the degree of conformance against a set of requirements, coinciding with Chrisman (1983) who described quality as " fitness for use ".

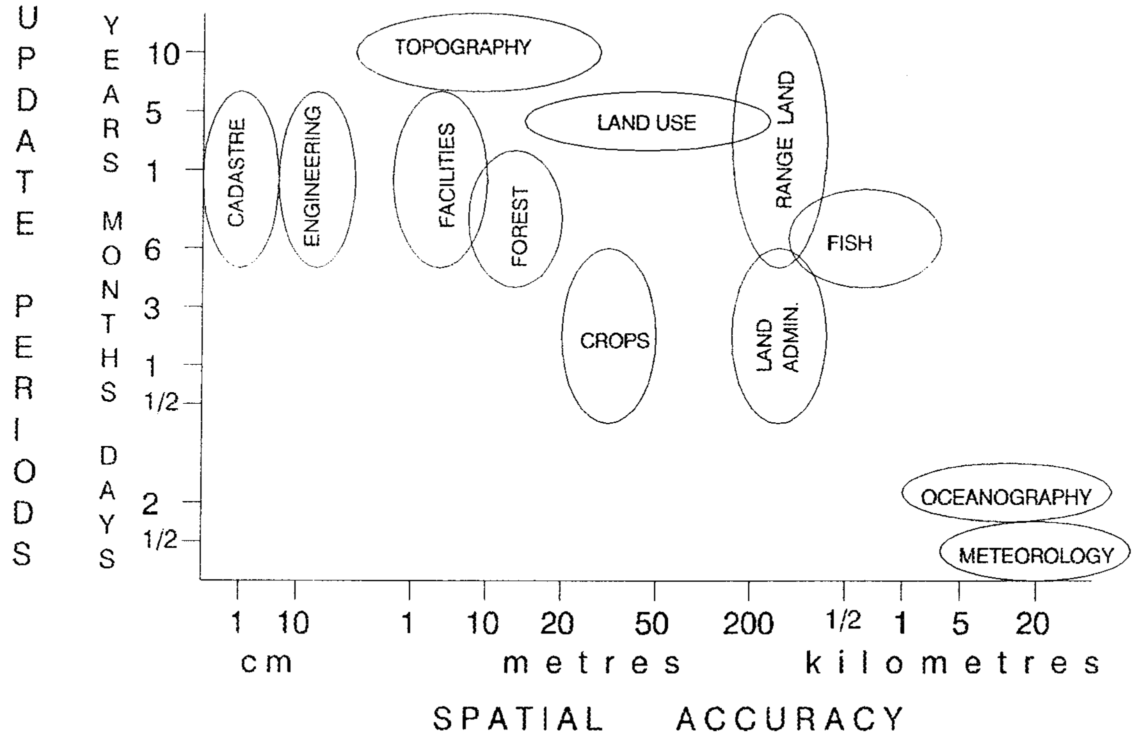


Figure 1. The Domains
(Source: Lyons and Sharma, University of Queensland, 1990)

THE PRODUCER'S GOALS

AUSLIG's primary objective regarding data production and conversion is to generate products that respond to the user community's demand. In order to be commercially viable, a product must support the widest range of uses and users and must accurately communicate to the user every detail concerning the data (truth in labelling), providing the basis to judge its fitness for a specific purpose.
A second but not less important objective is to generate a homogeneous product while running multiple production lines. This presents many challenges when the product is a representation of complex situations like topographic mapping, where the complexity of the real world makes every area a unique mini-world and therefore, it is virtually impossible to design a "one push button" mapping system.

In order to achieve homogeneity, a technical specification must be developed and strictly followed by the production lines. From this point of view, quality is measured against the technical specification, which is in fact a compendium of the producer's requirements.

A third objective is to reduce to a minimum the indirect cost of data to the user. This is the cost of cleaning the data after being purchased, cost that can easily exceed the actual purchase cost.

THE COST OF QUALITY

Improving the accuracy, adding new detail, adding more attributes, etc. are different ways of improving the quality of data. But when a technical specification defines the product, these issues are replaced by others related to the compliance with the specification.

No industrial process will produce a perfect product at all times. The question is then, how good should the product be made?. Considering this, the technical specification must also reflect the feasibility of the requirements, knowing that smaller error margins imply higher production costs.

Another aspect of the cost of quality is the cost of assessing the quality of the data. As expressed by Aronoff (1989) "The cost of assessing data quality varies with the degree of rigour required. The more rigorous the data quality testing, the more costly it becomes. This is not only a result of performing the test, but also of the delays caused in the production process to perform the tests and correct errors. For this reason, the level of testing should be balanced against the cost of the consequences of less accurate data or a less rigorously confirmed level of quality."

The cost of assessing data can be minimised if the technical specification and the production procedures are properly designed, e.g.: knowing that the planimetrical accuracy will be determined primarily by the scale of the source material (if converting analog data), the production procedures must be designed so that the component of the total error they generate is negligible compared with the source material component.

QA IN THE GEODATA TOPO-250K PRODUCTION LINES

GEODATA TOPO-250K is a topographic database that will cover the whole country. It is produced in tile units and sold by themes (every tile covers an area equivalent to a 1:250000 NTMS map and is broken into three themes: Hydrography, Infrastructure and Relief).

The main features of this product are:

1. The traditional hydrography is separated into a linear and a polygon layer.
2. The gaps left by water bodies in the streams network are filled with connectors, thus ensuring full connectivity.
3. Bridges. tunnels and railway stations are user nodes in their respective layers.
4. Vehicular ferry routes establishing full connectivity in the roads network.
5. Edge matching of graphics and attributes.
6. Unique feature identifier (UFI) system implemented to provide incremental updates.
7. Feature specific quality information.
8. Tile specific quality information.

These features make TOPO-250K the first of its kind worldwide. The complexity of the tasks required to meet the technical specification are multiplied by the mass production needs and in this environment, the result of every process depends on the previous one, making the implementation of QA a key issue for the project.

In practical terms, the implementation of QA in the production lines translates in:

- Extensive R&D for processes design.
- Design of effective error detection tools.
- Writing detailed documentation (work instructions and procedures) for the operational staff.
- Training of operational staff in the work instructions.
- Ensure that all production lines follow the procedures and use the same software.

While dealing with complex scenarios like the production of structured topographic data, R&D is the stone that guarantees solid foundations for the production processes. It provides for every potential situation that could arise later.

This is important for all processes but fundamental for the more complex or novel ones, that in the TOPO-250K production, are the separation of the lines and polygons in the hydrography, edge matching, management of the UFI and the quality attributes, etc.

The processes must be thoroughly tested and tuned up to provide the desired output under any variation from the theoretical conditions assumed in the initial development. The more testing is carried on, the less modifications will be needed once production started, saving time and resources.

Another important requisite is the design of recovery methodology that will care for situations like system crashes or process failure due to unexpected conditions of the data, not envisaged during the design stage, or human error.

The results of any process are usually dependent on the results of previous ones, therefore they must be continuously monitored by error detection tools that should also be thoroughly tested. These tools can be automated, semi-automated or manual and may include single, double and cross checks, plots, selections, screen panning, etc.

During the production of an average TOPO-250K tile, an operator will check at least 35 log files, 12 reports and 8 plots. Another 47 fully automated checks and 30 procedural or embedded checks are also carried out.

The documentation developed for the production staff must represent accurately and in full detail the production methodology designed. It must ensure the successful production of the tile despite any external problems involving staff, hardware, etc.

The operational staff must be trained in the procedures. As human intervention is one of the main sources of error, this training must focus around the interdependency of processes, their rationale and their contribution to the final product.

QUALITY CONTROL: THE LAST STEP IN THE QA PROCESS

Once a database is finalised it is submitted for Quality Control (QC) to an independent group: the Validation and Testing Section.

QC is a final and exhaustive inspection of the data before archival. As the control of every aspect of every feature is virtually impossible to be achieved, AUSLIG contracted the services of Dr. Des Nichols, Dean of the Faculty of Economics and Commerce, Australian National University, to investigate the best ways of implementing the QC tests in a statistical manner to ensure the compliance with the technical specification.

This consultancy established that from the 90 tests approximately 50 had to be carried out at all times (compulsory tests) and the remaining tests could be selectively carried out as they had a consistent history of passes. From this last category (spot tests), 2 randomly chosen tests are performed on the data as indicated on figure 2. This classification is revised fortnightly considering the latest performance results of every test.

The tests can also be classified according to their suitability for sampling in full population tests and sample tests. The sample tests are subdivided into two subclasses: when the full population is known and when it can only be estimated.

The first of these subclasses uses the Australian Standard AS1189-1989 "Sampling Procedures and Tables for Inspection by Attributes" in a straightforward fashion for the selection of sample sizes. For the second subclass the population is estimated as dense, medium or light. This translates into the division of the possible population numbers into three ranges and the higher sample size for the selected range is adopted from the Standard, consequently, these tests are always oversampled. In this way, AUSLIG ensures that the specification is met with 99% confidence.

The numerous checks carried out during production and the QC tests generate information about remaining problems. This information is analysed by AUSLIG's applications developers in charge of the maintenance and tuning of the processes, and the necessary corrective actions are implemented. In this aspect QC is another step in the QA process.

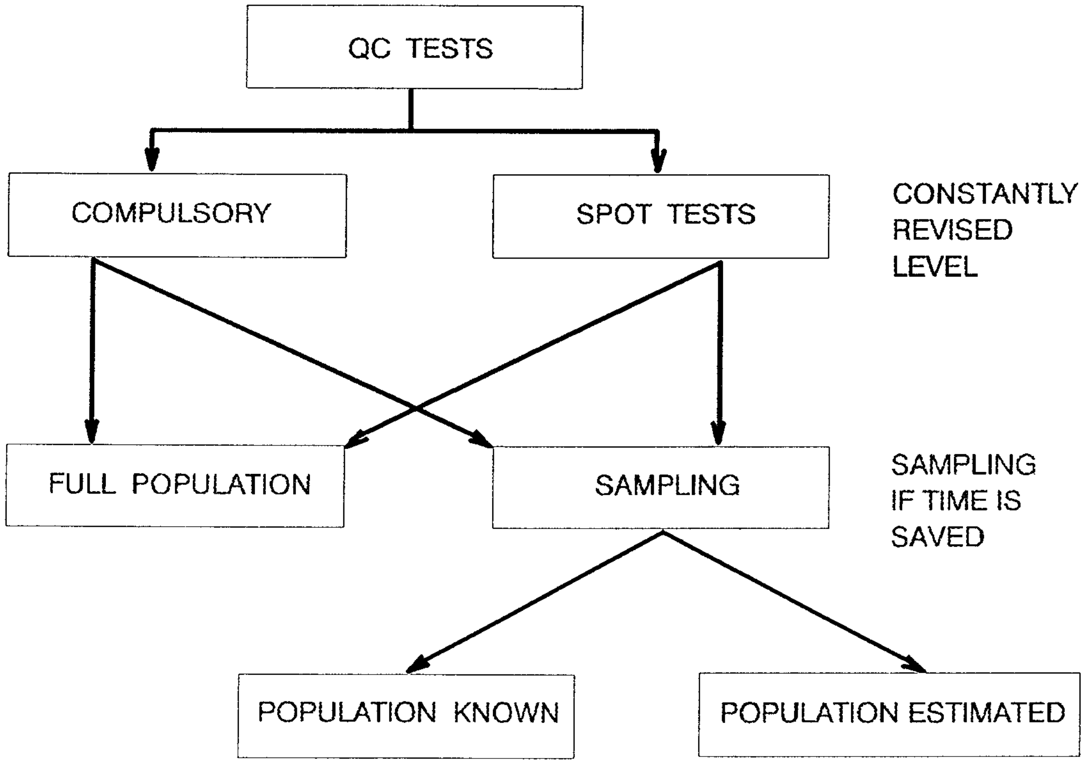


Figure 2. QC Tests Classification

TELLING THE USER

After ensuring that the specification standard is achieved, the final step is to transmit all relevant information to the user. This is done in four different ways:

1. Data User Guide: it is a 105 page document containing general information about digital vector data, essential characteristics of TOPO-250K, detailed structure and contents of every theme, available transfer formats and media, and a data dictionary.

2. Application Format Guide: it is a document describing format specific conventions, data loading operations, theme structure particularities for the format and format specific characteristics of the data.

3. Data Quality Statement (DQS): Anticipating the adoption of SDTS as an Australian standard, the DQS is an ASCII file containing the following information:

a. Tile quality information: it consists of metadata specific to the tile covered by the data file. It includes an exclusion of liability, conditions of use of the data and information about identification, formats, coordinate system, features in the file, WGS84 conversion parameters, point density, selective revision and edge matching.

b. Product quality information: it comprises five sections named Lineage, Positional Accuracy, Attribute Accuracy, Logical Consistency and Completeness.

An example of the DQS is included in the Data User Guide as an appendix.

4. Feature Specific Quality Information: at the lower level, every feature in a TOPO-250K database has four quality attributes associating basic metadata to the features:

Feature Reliability
Attribute Reliability
Planimetric Accuracy
Elevation Accuracy

These attributes are the basis for the evaluation of the accuracy and reliability of the results of GIS operations carried out by the users.

CONCLUSION

A QA system is an effective way of simultaneously achieving efficiency and excellence in the production environment.

As the land information industry moves towards TQM, there will be better and more.consistent products available, developed in consultation with the user community, as it is in TOPO-250K.

The old concept of equating quality with spatial accuracy is no longer satisfactory. Nowadays the quality of data can only be established if all the metadata is available to the user.

Although SDTS will provide guidelines for the transfer of metadata, the generation and communication of metadata files and attributes can be implemented today. The TOPO-250K experience is a successful example of this early implementation.

REFERENCES

Aronoff, S. (1989). "Geographic Information Systems: A Management Perspective", WDL Publications, Ottawa.

Chrisman, N. (1983). "The Role of Quality Information in the Long-Term Functioning of a GIS", Proceedings of AUTOCARTO6, Vol. 2 ASPRS, Falls Church.

Maguire, D., Goodchild, M., Rhind, D. (1991). "Geographical Information Systems: Principles and Applications", Longman Scientific and Technical, Essex.

Australian Key Centre in Land Information Studies (1990). "AKCLIS GIS Training Program Documentation", AKCLIS, Brisbane.

AUSLIG DAS, (1993). "1:250000 Topographic Vector Product Production Procedures, Section 8", AUSLIG DAS, Dandenong.

AUSLIG DAS, (1992). "H250TS V3.0 and I250TS V3.0 Technical Specifications", AUSLIG DAS, Canberra.

AUSLIG DAS, (1992). "TOPO-250K Data User Guide", AUSLIG DAS, Canberra.

An Experiment with Automated Acquisition of Classification heuristics from GIS

P.W. Eklund and A. Salim
Department of Computer Science
University of Adelaide
Adelaide 5005
Australia
email peter@cs.adelaide.edu.au,
ahmed@cs.adelaide.edu.au

Abstract

This paper explores the potential for symbolic inductive learning in GIS [ZG92]. Given that GIS's contain large amounts of data, the potential to apply standard inductive machine learning techniques[Qui86, BK87] gives scope to automate classification rule acquisition in land management applications.

This paper reports on an experiment in the domain of salinity class assessment [Kir92] where heuristic rules were used to construct an expert classification system. Classified data is passed on to the inductive learning program and its task is to replicate the classification rules used in the expert system. The initial goal is to restructure the flat rule-base and affect efficiency improvements in the expert classification system.

1 Introduction

The techniques described in this paper are of interest for two reasons: (i) inductive learning may allow an expert system's rule base to be reorganized to improve the performance and efficiency of the expert classification system (a crucial factor in classifying databases of GIS magnitude); (ii) inductive learning can be used to "discover" new domain knowledge in the form of classifiers; and (iii) the technique can be used to generalise classification rules from sample datasets. Point (iii) implies that field studies which classify small parcels of land can be used to scale up classification studies. This paper reports on results for points (i) and (ii). We leave (iii) as future work.

The objective of any classifier is to determine a decision procedure which will allow entities to be assigned to categories or classes on the basis of attribute values. Inductive learning is a branch of symbolic learning which constructs a classification procedure (or decision tree) on the basis of examining sets of entity examples that are already classified.

A set T of training instances of n classes $C_1, \ldots, C_n$ is presented to the program. The learning procedure is applied recursively until the only training instances remaining are those of the same class. All the instances of this class classify to this a single *leaf node* in the decision tree. At this point no discrimination is necessary and the process halts. If the training set is empty at the point where a leaf node is created then the result is so called *null* node. This represents an attribute value combination for which there is no example in the training set.

Finally, if the training set contains a mixture of classes then a single attribute is chosen which discriminates in an optimal way between the remaining classes. An information theoretic metric (called *gain*) selects an attribute a_i which becomes an intermediate node in the decision tree. The procedure repeats for each training set partitioned by a_i's attribute values. These values become branches of the decision tree.

For example in fig 1, the gain criteria selects ground water depth as the dominant discriminator. The training set is partitioned into two sets, one for training instances with ground water depth ≤ 3 and the other for ground water depth > 3. Each of the partitioned training sets are considered for each subtree of the decision tree. In the subtree where ground water depth ≤ 3, slope is selected as the next node in the subtree. If the slope is > 3 then all training instances with these features are of class 3. This equates to a single classifier:

```
Rule1: If ground water =< 3 then
    if slope > 3 then
          class 3 (LPD)
```

This process is repreated, forming new training sets and selecting attributes for discriminating classes in the training set until all training instances classify to a single leaf node (which defines rules such as rule 1)[1].

[1]The inductive learning system used in this experiment is C4.5. This system is available at nominal cost to academia (see Quinlaln[Qui90] for details).

2 Methodology

2.1 Salinity Classification

In the prototype GIS, farming land is subdivided into a database consisting of 302 rows by 402 columns (an area of approx 10×12 sq. kms). Each cell represents an area of 30 square metres and is described by the following twelve (12) attribute/values in table 1. Each attribute represents a data layer in the GIS database.

Attribute	Type	Values
Ground water	discrete 14	$\{0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30\}$
Elevation	discrete 6	$\{430, 450, 470, 500, 530, 560\}$
Slope	discrete 8	$\{0.6, 1.3, 2, 3, 5, 6, 20 > 20\}$
Geology	discrete 8	$\{QrSpa, QrLad, Pft, Phal, Pbi, Plg, Pls\}$
Soil	discrete 13	$\{sk, sk/cl/, ysil, yrl, cl/ysil, cl/yrl, cl,$ $besicl/ys, alluvial, bsl, bcl, bl/bcl, bl\}$
TMBAND1	continuous	0.45-0.55μm
TMBAND2	continuous	0.52-0.60μm
TMBAND3	continuous	0.63-0.69μm
TMBAND4	continuous	0.76-0.90μm
TMBAND5	continuous	1.55-1.75μm
TMBAND6	continuous	2.08-2.35μm
TMBAND7	continuous	10.40-12.50μm

Table 1: GIS schema for prototype database, TMBAND1 - TMBAND7 are digital spectral ranges for the seven (7) LandsatTM bands.

This data has been used as input into an expert system[KK93] whose rule base was constructed by domain experts. The task of this expert system is to determine the salinity class of a given cell. For example, a high potential discharge (HPD) area may consist of a number of cells that are allocated a salinity class of six (6). Such a classification is determined by the a number of rules, one of which is the following:

```
RuleX: if soil type = {3,5,7,12,13} then
    if geology is class {1,2, 3,4,5,6,7} then
        if elevation =< 2 then
            if slope =< 3 then
                    salinity class 6 (HPD)
```

139 rules of this type are defined for all six (6) classes of salinity. These are summarized below.

Description	Class	#Rules
Existing discharge	ED (1)	20
High potential discharge	HPD (2)	20
Low potential discharge	LPD (3)	33
Low potential recharge	LPR (4)	37
Medium potential recharge	MPR (5)	19
High potential recharge	LPR (6)	10
Total		139

Table 2: Salinity classes, there corresponding descriptions, and rules numbers associated with each class.

Rules are used to iterate over the GIS in the usual way and they subsequently derive a salinity map of the database and its associated spatial extention. The results of the expert system (called SALTMANAGER) are presented in Kirkby[KK93]. During the process of developing the system the several inefficiencies where noted: (i) I/O blocking cell retrival and storage to the Oracle database from the expert system shell; (ii) block structuring the classification rules so that they where mutually exclusive; (iii) the structure of the individual expert system rules. This paper concentrates on points (ii) and (iii).

2.2 Improving Rule Structure

Production rules were acquired through interviewing domain experts and reflect general purpose heuristic knowledge concerning salinity assessment in land management. They are therefore readable to the domain experts and understandable when considered singularly. Each rule represents a symbolic representation of a logical "piece" of expert knowledge.

The idea is illustrated by the following rules:

```
if gnd water =< 3 then
    if soil type = {3,5,7,12,13} then
        if elevation < 3 then
            if geology is class {2, 3, 4} then
                cell is class 1 (HPR)
if gnd water =< 3 then
    if soil type = {3,5,7,12,13} then
        if elevation >= 3 then
            if geology is class {2, 3, 4} then
                cell is class 2 (MPR)
if gnd water =< 3 then
    if soil type = {3,5,7,12,13} then
        if elevation < 3 then
            if geology is class {7, 8, 9} then
```

```
                    cell is class 1 (HPR)
if gnd water =< 3 then
    if soil type = {3,5,7,12,13} then
        if elevation >= 3 then
             if geology is class {7, 8, 9} then
                 cell is class 2 (MPR)
             ........
             ........
```

It is not necessary to test alternative conditions when they are a natural consequence of a prior rule failing. The above rules can reformulated by block structuring and eliminating unnecessary conditional tests as follows:

```
if gnd water =< 3 then
    if elevation < 3 then
        if soil type = {3,5,7,12,13} then
            if geology is class {2, 3, 4} then
               cell is class 1 (HPR)
            if geology is class {7, 8, 9} then
               cell is class 2 (MPR)
        elseif geology is class {2, 3, 4} then
               cell is class 1 (HPR)
            if geology is class {7, 8, 9} then
               cell is class 2 (MPR)
               .......
               .......
```

Additionally, it is preferable to block structure rules in terms of the "context" of the classes they assert or refute - classes are mutually exclusive[2], e.g.

```
if not (class 1) then
    if not (class 2) then
        if not (class 3) then
            if not(class 4) then
                if not (class 5) then
                    class 6. {default class}
```

This means that the inference engine need only consider a conflict set[3] with a maximum

[2]If one rule succeeds there is never any need to consider other rules. We need a structure which provides the direct equivalent of the "cut fail" combination or Prolog [Bra90].

[3]The conflict set is the set of rules in a knowledge-based system which are competing for control of the inference engine. Any one of these rules is eligible to be "fired", and usually some criteria, such as relevancy, is established for judging which is chosen.

of 37 rules[4] rather than 139[5] at a time. Nor does the inference engine search need to search for further solutions once a single rule has fired.

Although this rule organization seems obvious to a computer scientist, it is unfair to expect GIS domain experts to confront such issues, particularly during the development cycle. Indeed, this absence of rule-base structure is the main argument for expert systems technology, i.e. that domain experts can construct their own expert system applications and modify them without being trained in computer science[Kid85]. Efficiency is seldom a problem in traditional expert system applications since input responses are typically keyed and the development benefits resulting from the "active expert" [San92] methodology outweigh efficiency disadvantages.

In GIS applications the emphasis on performance remerges with renewed vigour since databases containing hundreds of thousands of elements are normal. For this reason, rule structure issues will make a significant performance difference.

To conclude, the intention of inductive learning applied to GIS classification tasks are thus threefold:

1. to detect and eliminate unnecessary conditional tests;
2. to suggest strategies for rectifying inefficiencies in the rules structure, i.e automatically determine class exclusivity;
3. to discover rules which may be shown to the domain expert and explained in terms of previously undiscovered domain knowledge.

It seems clear that 1. and 2. are well within the realm of inductive learning. The likely inefficiencies that result from a flat rule-base structure iterating over hundreds of thousands of tuples, and the prospects for speed-ups from restructuring, are themselves justification for applying inductive learning. However, the prospect of "discovering" new and previously unknown classifiers from the dataset, and in so doing provide new classifiers to the original classification system, is of particular interest in this experiment.

2.3 Inductive learning and the rule-base

As seen from Table 2, the SALTMANAGER expert system consists of 139 rules in total. These rules where fired over each of the 12-tuple cells of the database to produce a salinity classification for each. This process was repeated for each of the 121,402 $\times$ 12-tuples to produce 121,402 $\times$ 13-tuple learning instances with Table 2's structure

[4] The maximum number of rules that determine any one class [class 4], see Table 2.

[5] The entire rule set.

and the addition of salinity class. This process takes several hours on the current configuration[6].

Once classified by the expert system this data was input to the inductive learning program C4.5[Qui90]. Fig 1 illustrates a partial view of the decision tree (approximately 1/8 of the total decision tree) generated by C4.5 from this data. This decision tree, once constructed, takes an order of magnitude less time than the expert system to classify the GIS database[7].

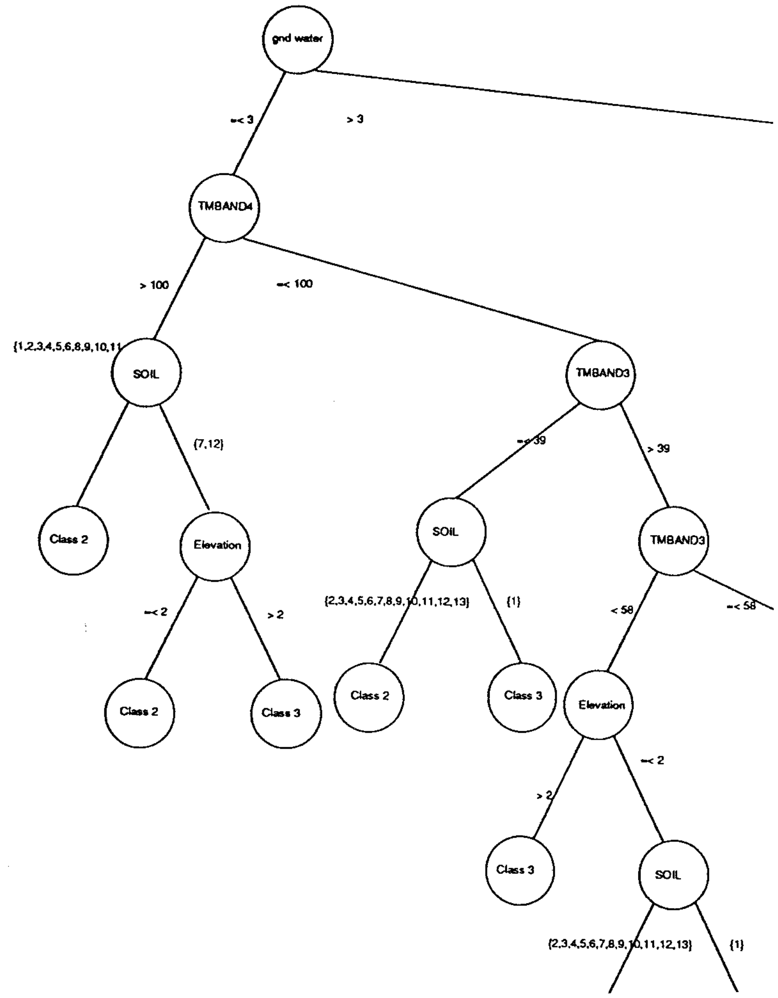


Fig 1: Partial view of salinity decision tree produced by C4.5.

The most obvious metric considers the number of leaves in the decision tree. This corresponds to the number of rules in the expert system knowledge base. A result summarized below in Table 3. This indicates that the inductive learning decision tree is structurally superior to the original rule-base. But what of its ability to classifying test instances correctly?

[6]Over 4 hours on a dedicated Sun IPX with 32 megabytes main memory.
[7]10 minutes on a dedicated Sun IPX with 16 megabytes main memory.

Description	Class	#Rules
Normal discharge	ED (1)	4
High potential discharge	HPD (2)	5
Low potential discharge	LPD (3)	3
Low potential recharge	LPR (4)	4
Medium potential recharge	MPR (5)	5
High potential recharge	LPR (6)	1
Total		22

Table 3: Salinity classes, their corresponding descriptions, and rule numbers associated with each class generated from C4.5.

In measuring the performance of this decision tree against alternatives we consider the usual pruning and simplification criteria[FI90]. The results are summarized below in Table 4.

Performance Indicators (initial&complete data)						
Gain Ratio	M1	M2	M3	M4	M5	M6
Before Pruning	4 (0.00016%)	48	36	84	505.75	1.75
After 25% simplification	4 (0.00016%)	38	30	68	638.84	1.79
After 10% further simplification	5 (0.00016%)	35	27	62	693.57	1.77
ID3 Gain	M1	M2	M3	M4	M5	M6
Before Pruning	5 (0.00021%)	48	36	84	505.73	1.75
After 25% simplification	5 (0.00021%)	44	35	79	551.70	1.80
After 10% further simplification	5 (0.00021%)	43	36	79	564.53	1.84

Table 4: Decision trees evaluation.

The results are a classification accuracy which indicates a high degree of confidence in the decision tree. The small % of errors (0.00016%) are more than accounted for by signal-to-noise on the satilite imaging and remote sensing.

2.4 Discoverying New Domain Knowledge

There are two issues concerning the nature of the inductive rules generated by C4.5 in this domain. The first is the logical structure and readability of the induced rules. Are they readable by the domain expert? Does the rule structure present the informationally most efficient means for determining a class at the expense of the domain logic? This concern was satisfied on the first run of the experiment. The domain expert could not only read the induced rules but also recognized his own knowledge: premise order may be slightly different but the attribute value tests were identical.

The second issue, and the most interesting aspect of this experiment, was to determine whether any new knowledge could be gleaned from the GIS database via inductive learning. By "new knowledge" we mean knowledge which was not explicitly used in the original classification task. The expert was surprised by the following combination of rules. Rules 1 and 16 appear to be in conflict with rule 14. The expert's conclusion was that this result, due to the illogical structure of these three rules in combination, suggested that the source remotely sensed data required reworking.

```
Rule 1:
     ground_water <= 3
     slope <= 3
     resimgb3as <= 39
     soil in {12, 13, 3, 5, 7}
     resimgb4as <= 100
->  class 2
Rule 16:
     elevation <= 2
     ground_water <= 3
     slope <= 3
     soil in {12, 13, 3, 5, 7}
     resimgb4as > 100
->  class 2
Rule 14:
     elevation <= 2
     ground_water <= 3
     slope <= 3
     soil in {12, 13, 3, 5, 7}
     resimgb3as > 58
->  class 2
```

On closer examination these rules 1 and 16 are consistent with HPD (class 2) which has yet to be come salt affected. TMBANDs 3&4 give indicators of healthy vegetation (low and high spectral responses in each respectively) which is experiencing flush growth: indicating a high groundwater table with low salt content. Although there is a high potential for salt affectedness it has yet to manifest itself. On the other hand, rule 14 indicates that the vegetation is stressed. This is a cell which has realised its high salt discharge potential. Goundwater is salt affected and the vegetation is stressed.

These two proper subclasses of HPD indicate that it is an aggregation of these two cases and suggests a new class.

Acknowledgement

The authors wish to thank S. Kirkby for his assistance and cooperation. This research was supported by the Australian Research Council.

3 Conclusion

The paper reports on an experiment in which classified GIS data is passed to an inductive learning program and its task is to replicate the classification rules used in the original classification. 80% of the training set was used to construct a decision tree and the remaining 20% as test data. Of the 24,280 test cases, 3 were incorrectly classified (error rate = 0.00016%). The test set was extracted from the database by ignoring every fifth tuple and partitioning it into the test dataset.

The rules suggested by the inductive learning program have yet to be recoded in the expert system shell, however it would not be premature to anticipate a substantial performance improvement.

Several interesting observations result from these experiments concerning the structure, efficiency, readability and content of the automatically induced rule-bases: (i) inductive learning allows a rule base to be reorganize to improve the performance and efficiency of the system; (ii) the induced rules correspond to the logical structure of the rules used in the original classification (iii) inductive learning can be used to "discover" new domain knowledge in the form of classifiers.

The work presented is preliminary but appears promising. Given that GIS's contain huge quantities of data (a necessary condition for successful inductive learning), the potential to apply symbolic machine learning techniques to classified data gives scope for automated the acquisition of classification rules for land management applications.

References

[BK87] I Bratko and I Kononenko. Learning diagnostic rules from incomplete and noisy data. In *Interactions in artificial intelligence and statistical methods*, pages 142–153. Technical Press, 1987.

[Bra90] I Bratko. *Prolog: Programming for Artificial Intelligence.* Addision-Wesley, 1990.

[FI90] U Fayyad and K Irani. What should be minimized in a decision tree? In *AAAI*, pages 749–754, 1990.

[Kid85] A L Kidd. Man-machine interface issues in the construction of and use of an expert system. *Man-machine studies*, 22:91–102, 1985.

[Kir92] S.D. Kirkby. Monitoring saline discharge areas farrell flat: A case study. *Australian Geographical Studies*, 31(1), 1992.

[KK93] S. D. Kirkby and F. Kurzel. Managing dryland salinisation with an integrated gis/es. In *Conf. on Land Information Management and GIS*, University of New South Wales, 1993.

[Qui86] J R Quinlan. Induction of decision trees. *Machine Learning*, 1:81–106, 1986.

[Qui90] J R Quinlan. *C4.5 Programs for Machine Learning*. Morgan Kaufmann, 1990.

[San92] K. Sandahl. *Developing Knowledge Management Systems with an Active Expert Methodology*. PhD thesis, Department of Computer and Information Science, University of Linköping, 1992.

[ZG92] K. Zhang and J. Giardino. Expert system based knowledge extraction from gis. In IGARSS, editor, *International geoscience and remote sensing symposium*, pages 319–321. IEEE, 1992.

PROVISION AND MANAGEMENT OF SPATIAL DATA IN NEW SOUTH WALES

Donald M Grant
Surveyor General of New South Wales
Land Information Centre
PO Box 143, Bathurst 2975
Australia

ABSTRACT

The Land Information Centre, Bathurst is the agency charged with taking New South Wales to the forefront of the global spatial information industry.

The Land Information Centre (the Centre) has seized this challenge by initiating three programs which will establish an advanced and successful spatial information industry within New South Wales. These projects encompass the control network for the State, a digital data base of the cadastral framework plus a natural resources inventory. Each program is described in this paper.

The Centre is responsible for the collection of a range of spatial data using either in-house resources, contracts to private business or purchase of existing data sets as well as the provision of this data in a useable digital format. By contracting work to the private sector the spatial information industry is fostered, in turn, prompting emerging needs to be identified and new products to be defined. The Centre has implemented education and market development strategies to create awareness of the potential of the spatial data and to promote its use. However, the successful utilisation of spatially referenced data also hinges upon the resolution of data access, pricing and maintenance issues to meet the changing needs of the users.

INTRODUCTION

There are innumerable varieties of spatial information. Spatial information is any data with a positional component, not simply the information pertaining to cadastral and engineering surveys. It is employed to support planning and decision making in a growing number of diverse applications from public safety to the sustainable development of a nation's resources. The spatial information industry has enormous breadth involving all aspects of data collection, storage, transfer, manipulation and analysis for an almost endless list of potential users and applications as well as research, training and technological capabilities.

The Surveyor General of New South Wales, through the vehicle of the Land Information Centre (the Centre), is tasked with the provision of survey control and the mapping of the State. The wider responsibility of the organisation is to satisfy public good obligations in a manner consistent with current government policy. This involves equipping the Centre with the latest technology and expert staff so that it can continue to supply high quality products and services ahead of public demand and provide a return on the State's investment in these resources. The modern focus of the Centre also effects technology transfer and a greater community awareness of the potential of spatial information, which inturn brings growth to the private sector.

Following a review of the State's surveying system in 1986, the Centre adopted a new focus - to convert, collect and supply spatial information in digital format. The trend towards computerised information was also gaining popularity in most other States and Territories of Australia, and worldwide. Within New South Wales the revolution towards digital data initiated three important programs. Firstly, the completion and upgrade of the State Control Survey to provide the fundamental framework to support a more cohesive cadastre and the development of a spatial information system. Secondly, collection of the Digital Cadastral Data Base, or DCDB, as the State endorsed base layer for use in spatial information systems. Thirdly, an ambitious project to collect and compile digital information layers for a Natural Resources Inventory (NRI) of New South Wales.

The benefit to the government of these three programs, in terms of revenue generated and expenditure saved, is consistent with the cost recovery policies of the New South Wales government and the growing commercial focus of the Centre. The strategies also offer a vital stimulus to the emerging spatial information industry in New South Wales and have created valuable intellectual property which can be marketed both in Australia and overseas.

The task of providing the spatial data collected through these programs to users and the procedures adopted by the Land Information Centre in the management and updating of the data are outlined in the next three sections of this report. Details of the purpose, goals, progress to date, maintenance procedures and use or expected use of the spatial data are given. The importance of the three programs to the development of the spatial information industry in New South Wales is also explained. The public safety applications of computerised spatial information provide an example of the Centre's role in the provision and management of land related data. Comments on the future of the spatial information industry in New South Wales conclude the paper.

SURVEY CONTROL NETWORK

In an ideal world, New South Wales would already be the proud owner of a complete and rigorous control framework as first advocated by Surveyor General Sir Thomas Mitchell in 1828. However, this utopia is only just nearing realisation. The Survey Coordination Act, 1949, assigns the Surveyor General with responsibility for the provision and management of the State Control Survey. The State Control Survey is comprised of the primary, secondary, tertiary and vertical survey control networks which provide the survey control data for all land parcels in the State. The State Control Survey program establishes, extends, maintains and promotes the use of each of these networks. The long term objective of the program is to achieve a complete, fully integrated cadastre which will minimise duplication of surveys, maximise use of existing surveys and provide a superior positional base for various spatial information systems.

To achieve this requires field survey and network adjustment duties as well as maintenance of a central register of State Control Survey information. This central register has been developed into a computerised Survey Control Information Management System (SCIMS) which offers timely access to reliable control data.

Programs to extend and upgrade the State Control Survey of New South Wales are designed to maximise the extension of the control networks, particularly in developing areas and major growth centres, enabling the declaration of these districts as Proclaimed Survey Areas. The private sector contributes to the control network when performing a subdivision by placing new survey marks and connecting the survey to these marks and the existing control to meet the requirements of the Survey Practice Regulation, 1990. This results in extension of the control network as part of the development process. In other initiatives, collaborate projects between the Centre's Survey Control Branch and the private sector are proceeding which will result in the provision of cadastral coordinates by the subdividing surveyor. The Survey Control Branch is also investigating legislative changes to permit greater effectiveness and efficiency in the extension and use of the State Control Survey.

Considerable progress in the State Control Survey program has already been made. The geodetic network for New South Wales is essentially complete and the State Control Survey now provides the fundamental framework for the digital cadastral data base. There are over 120,000 State Control Survey Marks and over 5,000 trigonometrical stations in use, placed strategically in our cities, towns, villages and throughout the rural areas of New South Wales.

As more permanent mark values have become available for use, enquiries through SCIMS have steadily increased with a marked increase in activity, particularly for horizontal control information, following the introduction of the Survey Practice Regulation, 1990. Implementation of the new Regulation is also responsible for the recent growth in the number of survey marks entered into the SCIMS data base as illustrated in Figure 1. The increase in the placement of new survey marks can also be attributed to the adoption of total station technology and adjustment software which has facilitated control surveys. Global positioning systems are the latest technology to impact upon control surveys carried out by staff of the Centre. In an eight week period last year, approximately 100 primary control stations and another 60 permanent marks were placed in the Western Lands Division of New South Wales. Now every settlement with a population of 1000 or more over this Western Division - being 42% of the State - has some form of geodetic control within the town or settlement itself.

Inspections by Survey Control Branch staff of areas where intensive control surveys are due to take place have revealed a survival rate of permanent marks comprising

STATE CONTROL SURVEY

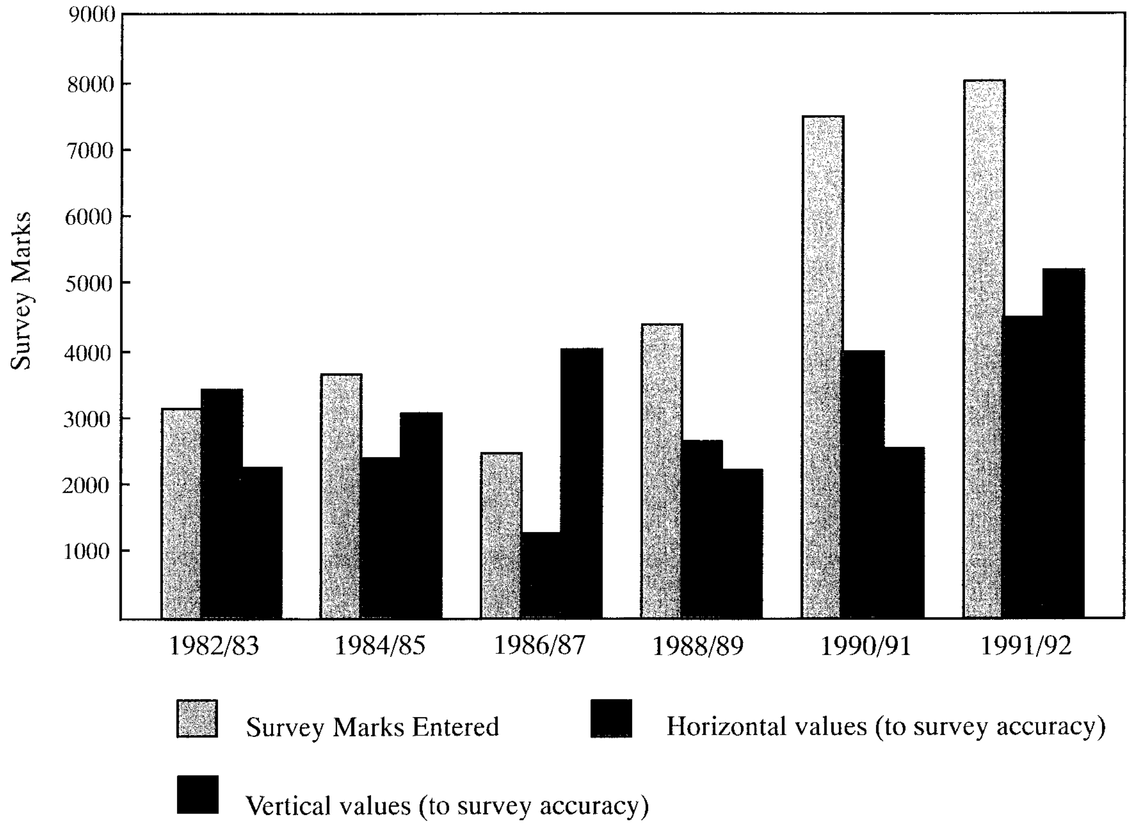


Figure 1: Survey Mark Progress 1982/92

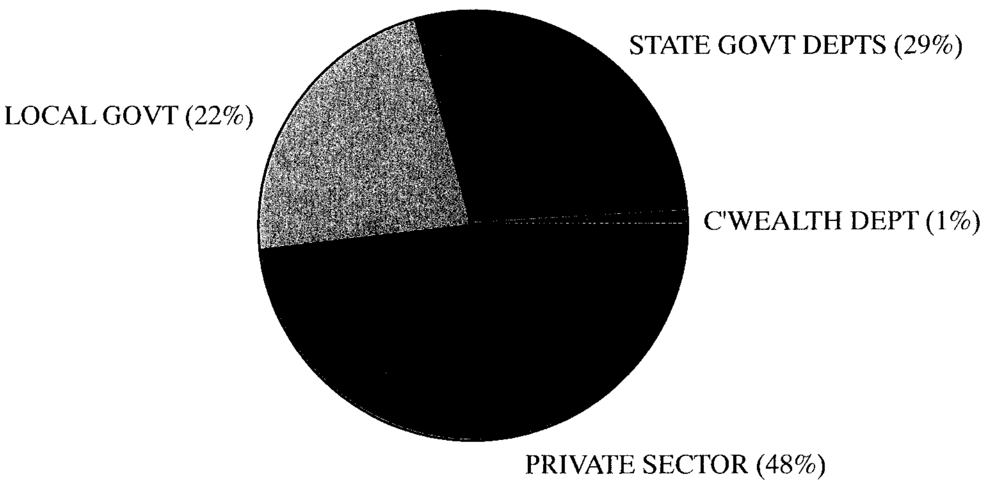


Figure 2: Users of SCIMS Enquiry Service

CALM LAND INFORMATION CENTRE

the State Control Survey of between 95 and 100 per cent. The permanent marks are protected under the Survey Coordination Act which requires local councils to preserve survey control marks. In addition, all surveyors are charged with reporting damaged or missing marks to the Survey Control Branch so that they can be replaced.

Survey control data is fundamental to all aspects of land development and information storage. The following examples indicate the variety of activities for which survey control is necessary.

- Residential and industrial development;
- Major engineering projects;
- Provision of communication services and other utilities;
- Land information systems;
- Mapping;
- Identification of flood prone areas;
- Mineral exploration and extraction.

Users of the system include all manner of government departments and authorities, city, municipal and shire councils as well as numerous private businesses. The proportion of each of these groups of users is shown in Figure 2. Information on survey control marks is obtained by these users to enable them to meet their obligations under the Survey Coordination Act, 1949 and the Survey Practice Regulation of the Surveyors Act, 1929.

At the present time, the search for survey control information is initiated by enquiries made over the counter, in writing, by facsimile and over the telephone to the Survey Control Branch. The required information is obtained from the SCIMS data base and by manual searches for the relevant diagrams and sketches. A consultant is currently developing a system to fully automate this process to enable the permanent mark number, coordinate value and relevant standard of accuracy to be produced and automatically dispatched to the user along with a diagram of the relative location of the marks and a sketch plan of each permanent mark. Testing of this system is scheduled to begin in a few months. Following the successful implementation of this fully automated process a program to make the service available on-line will commence.

Remote access to part of the SCIMS data base is soon to realised through the development of a Bulletin Board service at Bathurst. Subject to further testing, the Bulletin Board will be fully operational within six weeks. Coordinate and height values will be available via a link by telephone modem.

DIGITAL CADASTRAL DATA BASE

The Digital Cadastral Data Base (or DCDB) is a computerised data base of the cadastral fabric of the State. It can be considered a graphical store of the land parcels, public ways, reserves, railways, administrative boundaries and relevant mapping easements together with Lot Deposited Plan numbers as unique identifiers. The DCDB is digitally collected from cadastral maps and existing digital data available at the time of capture. This information is stored for all related parcels in a single data base of what has proved to be the most complex cadastre in Australia.

The purpose of the DCDB is to:

- Create administrative efficiency and eliminate the duplication of equipment, staff, expertise and time;
- Improve the availability of land related data;
- Facilitate public access to information;
- Enable the development of new, high quality land information products;
- Facilitate effective management of the land resource by providing the foundation for spatially referenced data in an interactive data base system;
- Assist all levels of government, public sector agencies and the business community to make better management decisions;
- Generate revenue for government.

The Digital Cadastral Data Base, or DCDB, project began in earnest following the approval of funding by government to accelerate the capture of the data base in 1991. At the beginning of April 1993, 3.2 million legal parcels or 4.1 million polygons have been digitised into the system. Capture of the total number of polygons, roughly five million, is estimated to be achieved by the end of 1993. It is anticipated that the storage requirement for the complete data set will be approximately five gigabytes. Figure 3 illustrates the progress made towards capture of the DCDB to the 1st June 1993.

Initially the DCDB was captured using the in-house resources of the Cadastral Capture Unit of the Centre but following the acceleration of the program contracts for the data capture were let to private companies. The role of the Centre changed, it became the coordinator of these contracts, responsible for checking and updating the digitised data. The New South Wales government encourages the involvement of the private sector in such projects. Indeed, the DCDB project was responsible for fostering seven private sector companies and stimulating growth in the emerging spatial information industry.

The DCDB is verified using an automatic process which compares the unique identifiers attached to the DCDB parcels against the parcel tags in the Land Titles Office's Automated Land Titling System (ALTS) and the Rationalisation of Property IDentification (RAPID) system. Manual updates of the DCDB are completed from new deposited plans sent from the Land Titles Office in Sydney to the Cadastral Capture Unit in Bathurst. The new plan is inserted and the surrounding DCDB parcels fitted accordingly, rendering the new parcel of higher accuracy than the existing DCDB data. An automatic process, operating along similar lines is currently being developed.

The digitisation, verification and update processes have been reviewed using Total Quality Management principles resulting in the improvement and streamlining of operations and significant savings in time and money.

NSW DIGITAL CADASTRAL CAPTURE PROGRAM

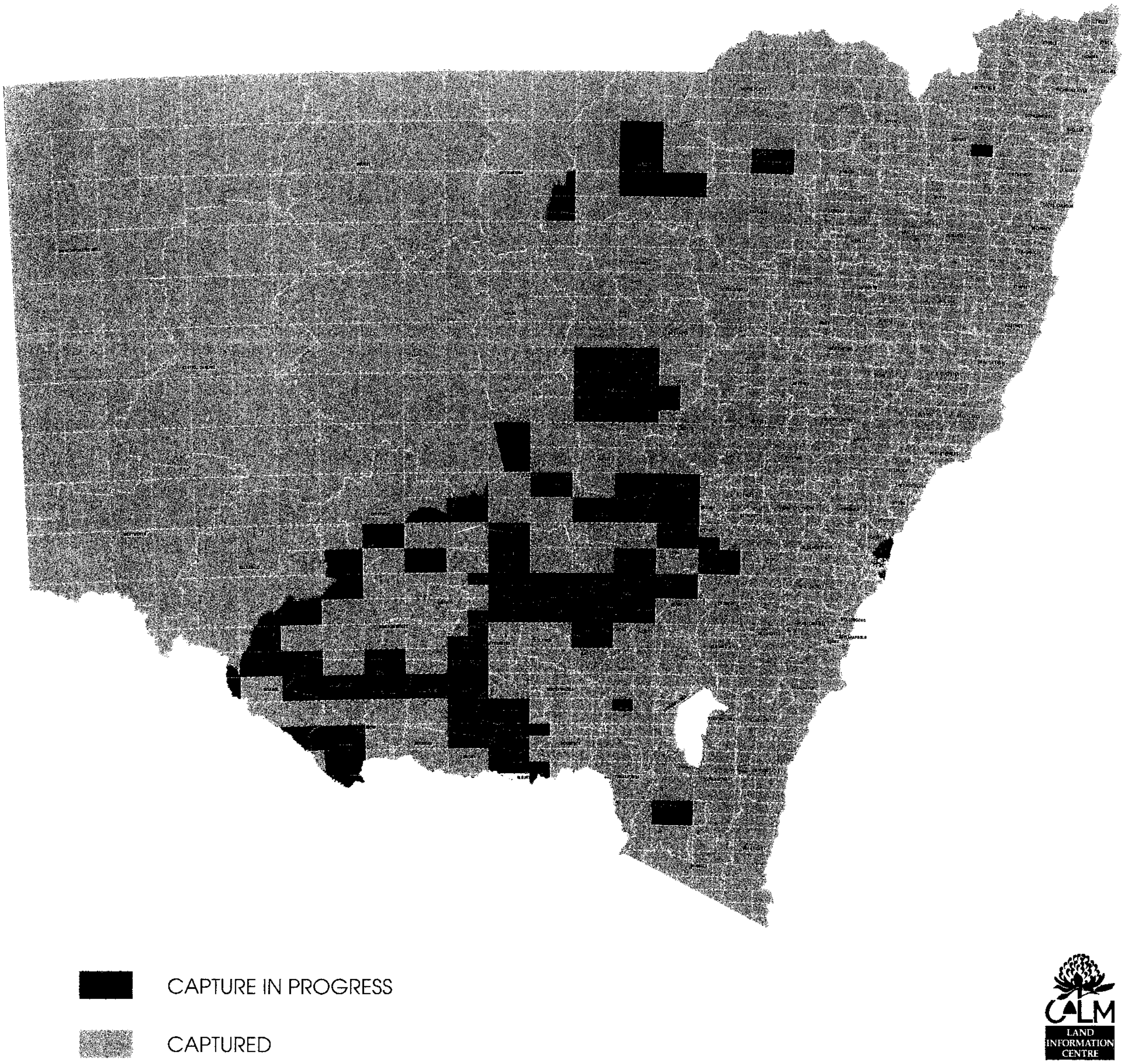


Figure 3: Progress of the Digital Cadastral Data Base as at 1-6-1993

The spatial integrity of the DCDB reflects the source material used for its capture and this accuracy has proved sufficient for the applications for which it was designed. However, to continue to meet the demands of users, programs to upgrade the spatial integrity of the data base have been investigated by the Cadastral Development Unit at the Centre. Interviews with government departments, local councils, surveyors and utility groups have revealed the need to upgrade the DCDB to a consistent accuracy of about ± 0.3 metres on urban cadastre and ± 1.0 metres on rural cadastre to provide the required cadastral coordinate information for location purposes.

The potential uses of the DCDB as a base layer in conjunction with other associated data are almost limitless and include applications in:

- Surveying eg. to locate survey marks near a specific parcel;
- Agricultural land management;
- Land use planning eg. soil erosion studies;
- Education eg. to determine school site location and school catchment areas;
- Utilities placement and maintenance;
- Local government rating, planning and administrative functions and to monitor change in and demand for resources;
- Transport for planning and maintenance of facilities;
- Natural disaster planning, prevention, containment and clean-up;
- Real estate, property development and investment;
- Provision of emergency services;
- Retailing eg. in marketing, planning shopping centres and computing distribution networks.

In addition, a customising service is provided by the Land Information Centre to generate products from the DCDB or to integrate the DCDB with a variety of other data bases to clients' specifications.

The soon to be completed digital cadastral data base has revolutionised spatial information in New South Wales. The DCDB will provide the fundamental spatial framework for the emerging Natural Resources Inventory. Perhaps the digital cadastral data base should be considered the first step in a giant staircase leading to the spatial information age.

NATURAL RESOURCES INVENTORY

Integral to the development of a Natural Resources Inventory (NRI) for New South Wales is the Land Information Centre's Primary Spatial Data Base (PSD) project. The PSD will provide the base level data upon which the NRI and other specific data sets can rest. When fully developed the PSD will include three dimensional data that will allow better modelling of critical environmental processes. The project involves the digital capture of five fundamental data layers: a digital surface model; drainage and shoreline data; the transportation network; baseline satellite imagery; and the major boundary types. The Centre has begun collection of the PSD using existing in-house resources and hopes to complete the project within two years given appropriate funding. The data layers are being digitised from the existing topographic map series at scales of 1:25,000 and smaller.

The PSD and NRI programs have been developed in parallel with changing government policy which is responding to an increasing awareness within the community of the need for sound management of our natural resources. The significance of the NRI has been recognised by the New South Wales government as recently as the 20th May, 1993 by the announcement of the Premier that a Natural Resources Audit Council will be created to "conduct an audit of the state's publicly owned land and natural resources." In addition, it was announced that five million dollars would be made available this financial year to accelerate the collection and coordination of natural resource information.

The NRI is essentially an industry wide program for the collection and maintenance of layers of natural resource information collected to common standards and capable of being integrated with other data layers. The purpose of the NRI is to provide a definitive, integrated and easily accessible information base to support decision-making by government, industry and the community on natural resource and environmental issues. The NRI will also:

- Reduce cost to government by eliminating duplication and the sharing of resources between agencies involved in the collection of natural resource information;
- Improve the quality of decisions to do with natural resources;
- Provide a fundamental information resource for policy formulation;
- Improve coordination between public sector agencies resulting in more efficient operations.

The Centre will contribute the PSD to the NRI and has taken on the role of agent responsible for coordinating activity for the NRI project. The Centre is also a source of technical support for the integration of the data and operation of the NRI.

At the present time, many aspects of the NRI are still in the conceptual stage. An application for funding to facilitate the coordination of the project and to accelerate its capture is to be resubmitted to Treasury in the near future. Some of the issues awaiting clarification are standards for data collection, mechanisms for accessing the information and the price of the data. A principle of the NRI is that it should be useable in a national context to complement data collected by other jurisdictions. To this end national data and transfer standards will be used where available.

Small working groups were convened in May 1992 to develop, layer by layer, the types of spatial and attribute data which would be useful to a broad range of users and to investigate the availability of such data. The layers of NRI data envisaged and the components of the PSD are shown in Figure 4. These data layers have evolved from a requirement identified at a national level through the Australian Land Information Council (ALIC), now ANZLIC, in early 1990. The NRI layers are at various levels of digital capture with some data sets currently unavailable in computerised form. The data sets are being collected by the relevant sectors of industry, including a number of government departments and agencies, as well as some involvement by academia.

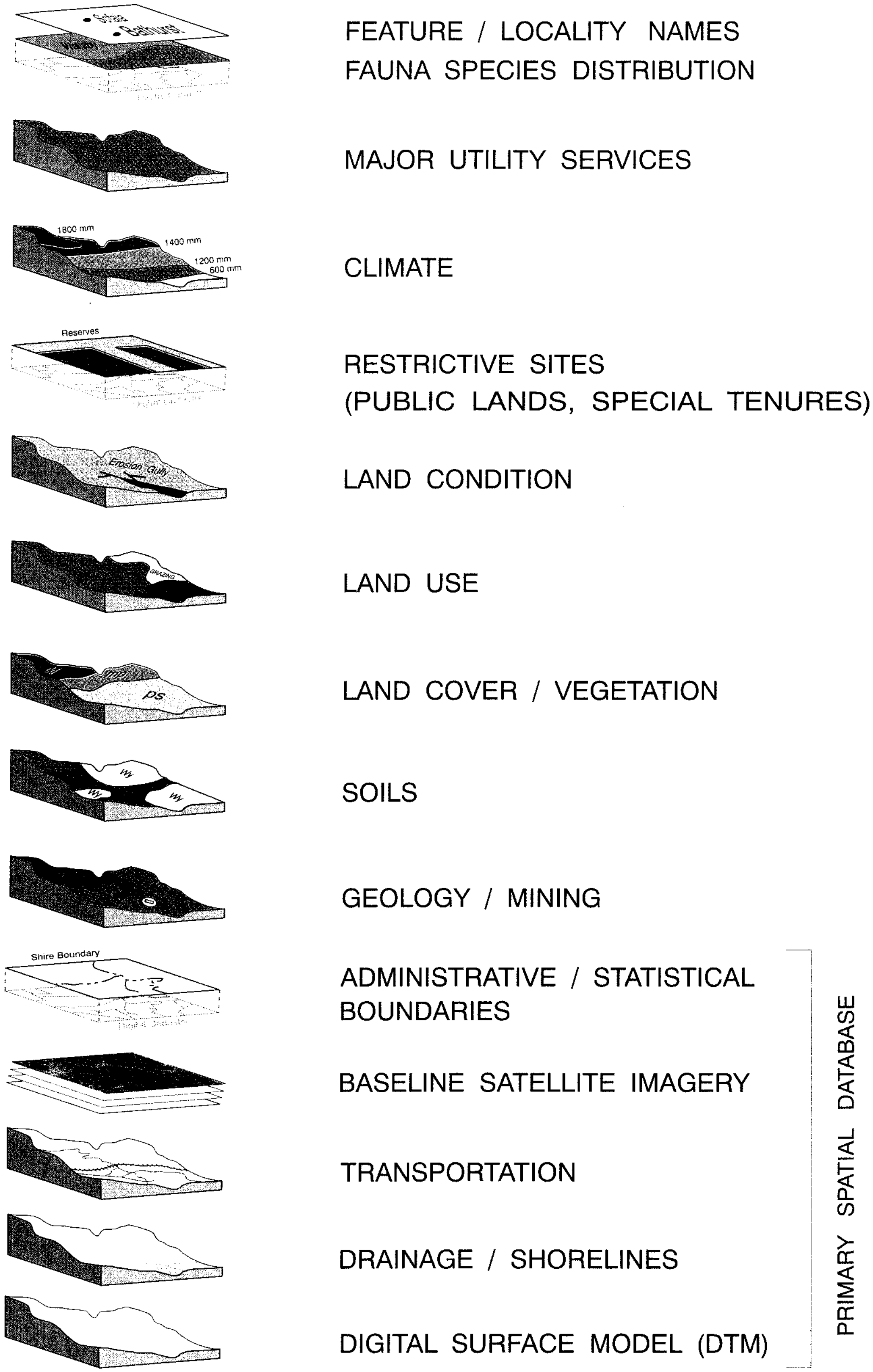


Figure 4: Layers of the Natural Resources Inventory and the Primary Spatial Database

CALM
LAND INFORMATION CENTRE

These agencies will contribute the data to the NRI but will remain trustees of the data sets. The NRI will be a dynamic data source, in that existing data is both updated and upgraded in quality over time, and if required by users, expanded in scope.

Further work must be done in the design of data layers to enable the diverse data sets to "fit" together so that users can combine data sets with some certainty that the results will be meaningful. A workshop, held in Bathurst in July 1992, concluded that impediments in the way of implementation of the NRI included lack of a data dictionary covering the intended entities contained in each data set, a diversity of co-ordinate and attribute classification systems amongst data sets, lack of complete coverage for some data sets and lack of a comprehensive data transfer standard. All of these obstacles can be overcome and the success of the project ensured through sponsorship at the political level, the provision of suitable funding and receipt of operational support from the government agencies involved in collecting data on natural resources and the environment.

The integration of basic data sets in the NRI has application for addressing a wide range of issues for State-wide, regional or sub-regional studies, such as:

- public safety;
- regional planning;
- pollution monitoring;
- conservation of flora and fauna;
- land management.

Applications for computerised natural resource type data at a national level are demonstrated by two projects currently being undertaken at the Centre.

The first of these is a vegetation mapping project covering the entire Murray Darling Basin using satellite imagery. The Centre is responsible for managing the project and providing the imagery for New South Wales. The project is sponsored by the Murray Darling Basin Commission and comprises a formal inter-governmental partnership between the Commonwealth, New South Wales, the Australian Capital Territory, Victoria, South Australia and Queensland. The project will collate and collect data that describes the structural and floristic characteristics of the vegetation in a digital form. This information will be included as the vegetation layers of the New South Wales' NRI. The data will provide a good baseline with which to monitor change, predict future impacts and evaluate the effectiveness of policies and action to manage the Murray Darling Basin's natural resources.

In another initiative, the Land Information Centre, together with the public sector mapping agencies from each State, Territory and the Federal government, have won a contract, which was signed on the 9th June 1993, to supply the digital base map data for the 1996 Census of Population and Housing conducted by the Australian Bureau of Statistics. New South Wales, through the Land Information Centre, has provided the focus for the project and will be responsible for the management and integration of the data sets supplied by each jurisdiction. The use of digital map data will aid the planning of the Census and facilitate the collection and dissemination of census statistics for the 1996 and future Censuses. The creation of a base level of data covering Australia is a valuable product that can be used for a wide variety of purposes on a national scale.

PUBLIC SAFETY APPLICATIONS

The use of spatial information for public safety applications is an excellent example of the potential uses of the spatial data supplied and maintained by the Land Information Centre.

The Centre began an initiative to realise the potential of digital spatial information for public safety applications by convening a Public Safety Working Group in September 1992. The Working Group comprised representatives of the New South Wales Police Force, the Ambulance Service, the New South Wales Fire Brigade, the Bush Fire Service and the State Emergency Service. These organisations have a common mandate to save lives and protect property as well as a collective need for basic spatially referenced information to enable them to perform this role.

The Public Safety Working Group meets monthly to discuss the applications and availability of the fundamental spatial data common to all organisations as well as the additional, specialised data that is necessary to the operation of each, with a view to incorporating this data into a computerised spatial information system. For example, the Fire Brigade, like all of the emergency service groups, requires a map of the road network with the appropriate street addressing, but by linking this map to a data base indicating the presence of hazardous chemicals and their composition the information system becomes uniquely useful for fighting fires. The Centre acts as an impartial coordinator of systems development for the emergency service authorities, as most lack a detailed knowledge of the potential applications of digital spatial data and the resources required to implement a spatial information system.

The fundamental spatial data required by the emergency service group is already held digitally and maintained by the Centre on behalf of the State, or is subject to a digital capture program. Examples of the commonly required data sets are the road pattern, landscape and digital cadastral data base. The benefit to the emergency service organisations of using State supplied data is that it provides a common reference surface within the emergency service group and also with other organisations such as the water and power authorities with which they communicate. In addition, the State data will be supplied to each emergency service as part of the Centre's public good obligation. The community benefit from utilisation of State data will demonstrate the value of the data and justify the expenditure on its capture.

At the present time, each member of the Public Safety Working Group is operating some form of computer aided dispatch system. As each of these systems gain compatibility through the use of State held basic data layers the emergency service groups will be able to realise the following benefits:

- More efficient operations by reducing the time taken to respond to and contain an incident, resulting in saving lives and property;

- Facilitate decision-making in emergency situations where the response characteristics are similar;
- Automate many manual processes, enabling more effective use of resources;
- Eliminate duplication of basic incident records;
- Opportunity to rationalise spatial information delivery for the authorities by creating a single distribution station for common incident data.

In recognition of these benefits, the New South Wales Fire Brigade have made a mandatory component of their tender for a fully automatic computer aided dispatch system, the use of Land Information Centre data. The efforts of the Centre have also come to fruition by the alteration of priorities in the proposed use of spatial information systems for the Police Force and the Workcover Authority. The use and integration of spatial information will now come early in the development cycle of these and other systems evolving for public safety applications.

CONCLUSION AND DISCUSSION

From the preceding discussion it becomes clear that the future role of the Land Information Centre will be that of an information broker packaging, selling and distributing spatial data and marketing the intellectual property generated from spatial information projects through consultancies. The information commodity will consist of the spatial data gathered for the State Control Survey, the Digital Cadastral Data Base, the Primary Spatial Data Base, as well as the Natural Resources Inventory. A by-product of these programs is valuable intellectual know-how and experience.

The provision of fundamental base layers by the Centre provides the foundation upon which a hierarchy of other information layers which possess a spatial component can rest. This service is essential to the development of the spatial information industry in New South Wales and the effective, responsible management of our resources.

The spatial information industry is growing at a phenomenal rate world-wide. Between now and the year 2000, the global market for converting existing government hard copy maps, charts and engineering drawings to computer readable form is estimated to be $US90 billion. Another report claims that the Aid Agencies of the world will invest $US25 billion in land and marine information systems over the next decade. Development loan funds from the international funding agencies are estimated to be $US100 million per annum for training alone (Price Waterhouse Urwick, 1992). The Centre's target for New South Wales is to build up to a 0.5 per cent share of this market. That equates to $US45 million per year in data capture and $US500,000 in training. Further growth of the Centre's expertise in digital information technology and the definition of opportunities for relevant education packages will ensure the Centre wins its market share.

The Centre is in the enviable position of possessing intellectual property which can be shared through consultancies to provide a mix of staff, training, expertise and management for spatial information initiatives and projects. This intellectual capability has great potential to further contribute to the local, national and international spatial information industry.

The goal of the Centre's Business Development Unit is to maximise the return on the State's investment in the intellectual property contained within the Centre by marketing this commodity in the international and domestic market places. The overseas marketing has culminated in the establishment of a joint venture commercial operation in the Guangdong Province of the PR China and planning for the creation of a joint venture marketing company with the PR China National Bureau of Surveying and Mapping. The Unit is currently running a small format aerial photography project in Lao PDR, with the prospect of performing the mapping of Vientiane, the capital of Lao PDR. In addition, negotiations are in train to supply management and expertise for mapping and systems development projects in SR Vietnam, Myanmar and many organisations within Indonesia. The Business Development Unit is also responsible for establishing a commercial company to be known as Austinfo Pty Ltd which will facilitate all commercial activity at the Centre.

In another initiative, the Australian Institute of Spatial Information Sciences and Technology, or AISIST, has been established to take advantage of the opportunities for vocational training in Australia and overseas. AISIST involves a consortium of Australian Universities, the New South Wales TAFE Commission, the Australian Key Centre in Land Information Studies (AKLIS), the Information and Technology Division of the CSIRO and the Land Information Centre at Bathurst. As part of the Federal Government's Jobskills scheme, AISIST is currently conducting a regionally based training program which, when completed, will provide the spatial information industry with approximately 100 fully trained practitioners throughout the State. On the international front, the Institute provides the training component of the Lao PDR project, with negotiations in progress to run spatial information systems training programs for public sector organisations in both Indonesia and Thailand.

These state-wide and international initiatives appear very promising for the spatial information industry in New South Wales, yet continued success depends upon the resolution, at both federal and state level, of a number of policy matters. The definition of data transfer standards, privacy, access, liability and pricing issues as well as the level of commercial enterprise permitted by government agencies, will affect the form and development of the spatial information industry. The current lack of policy can largely be attributed to the absence of precedents, the very sensitive issues that must be addressed and the slow machinations of government.

The future of the spatial information industry lies with the collection and use of socio-economic data sets. The socio-economic layers consist of demographic and administrative information that has a relationship to specific locations. The Centre's programs will provide the foundation for the integration of these layers, creating an information system of truly awesome dimensions. Such a system has the potential to enhance our decision making capabilities but also possesses some negative implications. The likelihood of invasion of privacy will increase with the synthesis of socio-economic information, but perhaps privacy will lose its significance in the face of data integration capabilities. How this conflict is resolved forms a challenge for the future.

ACKNOWLEDGMENTS

My thanks to Brigitte Krogh and all of the helpful staff at the Centre who provided information and advice to assist with the preparation of this paper.

REFERENCES

Grant, D M (1993). "Leaping the Hurdles to Cadastral Reform in Australia", proc. SIRC Colloquium, Dunedin, New Zealand, May, 307-320.

Price Waterhouse Urwick (1992). "AISIST Spatial Information Systems Industry Needs Analysis", Sydney, March.

MOSAIX - A Spatial Data Integrator

or 'How to succeed at GIS without even crying'

Roy Hill B.Sc.(Hons), M.I.E.Aust, M.A.C.S
Technical Director
CADCOM Pty Ltd
1/170 Pacific Highway, Greenwich
NSW2065,Australia

ABSTRACT

In the corporate environment, spatial and aspatial data is available from many sources.The question often asked is 'How can I view the data together?' The answer varies dependent on the amount of translator technology available.This paper is a brief exposure to a local, Australian owned product, **mosaiX**, that combines the use of current server technology to solve this age-old problem. Data stays in its natural binary environment and is combined 'on the fly', whether from **multiple concurrent databases** (e.g. Ingres, Oracle, Informix), **raster** (e.g. Tiff, pcx) or **proprietary spatial formats** (e.g. ArcInfo, Autocad, Genasys, Geovision, Intergraph, Landmaster).

INTRODUCTION

Mosaix is a sophisticated object oriented application development tool designed to enable multiple technologies within a single language environment.

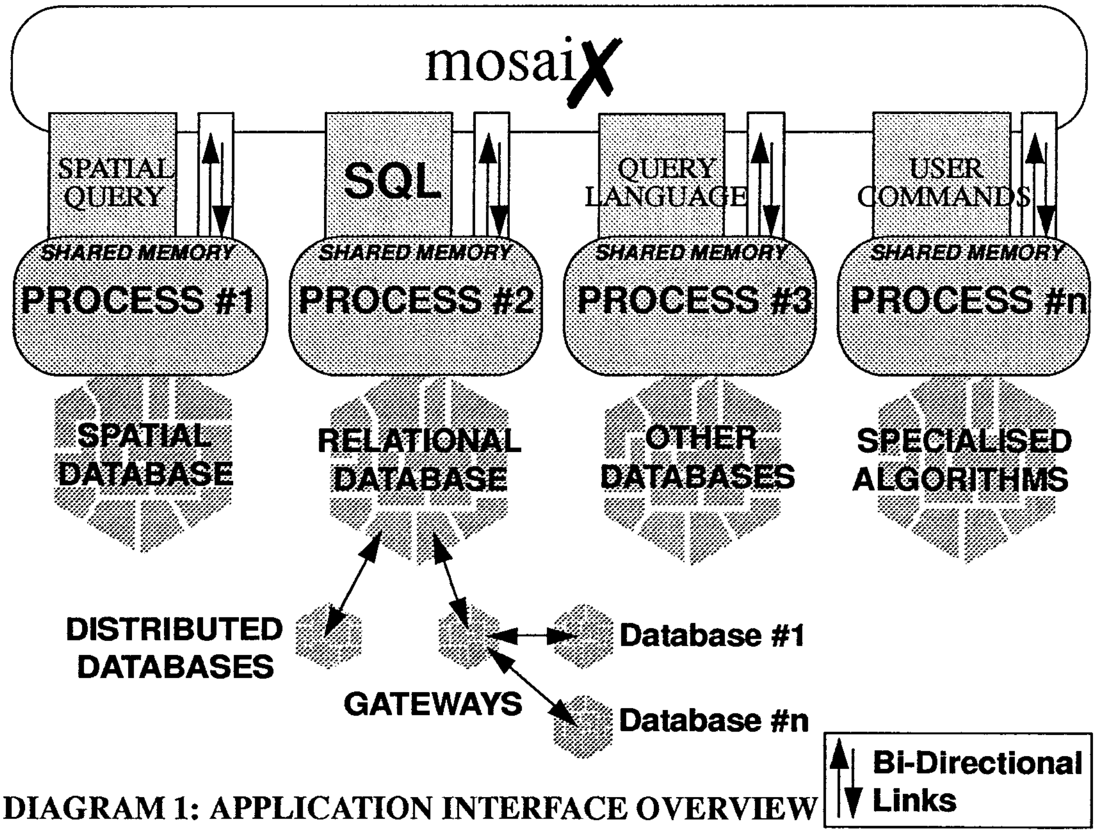


DIAGRAM 1: APPLICATION INTERFACE OVERVIEW

It is the ultimate client application environment by providing a high level language interface to standard technologies such as GUI,multiple concurrent database, 2d/3d Vector and Raster Graphics and eventing by internal and external processes whilst maintaining portability across differing operating systems,windowing systems and relational databases. The language itself can be likened to a 4GL as it supports all of the standard procedural language constructs (such as if then else, while, case) as well as the object oriented constructs such as user defined classes, inheritance,etc. Generic datatypes found within the language environment contain high level functionality which enables the user to quickly develop applications. An example of this is the window datatype which allows the user to open a 3d graphics environment with spatially co-ordinated vector/raster in a single command. As such,mosaiX is a natural tool for prototyping but mosaiX's strengths are not only in the rapid development of applications ,they also provide the finished product.

As new technologies are always emerging mosaiX has been designed to be easily extended through the mosaiX Application Program Interface(API). The mosaiX API provides an open interface which allows system integrators (and mosaix itself) to tightly couple specialised processes and/or database interfaces to the mosaiX language environment.(e.g. Windows 4GL,SCADA). As these external processes communicate commands and data through shared memory, optimum performance is obtained without sacrificing modularity and without any impact on the development and maintaince of mosaiX or other co-operating processes.

In most organisations, the next wave of applications require access to multiple database systems which have existed for many years and newer systems that are being introduced. Through the mosaiX language environment it is possible to leverage and integrate these systems with minimal impact on the organisation and gain maximum utilization of the data.

1.0 mosaiX Language Environment: EXAMPLE

Hypothetical macro problem:

1) Given **3 concurent databases** that exist on **'the Network'**
   - a) **Ingres** with Valuer General reports
   - b) **Oracle** with Land Title Information Reports
   - c) **Informix** with a **Map Management Inventory System** that contains tables that relate **the Network location** and **spatial extremities** of:
     1) **Land Parcel Polygons** in **LANDmaster** format **(*.mod)**
     2) **Road Centreline Linestrings** in **Autocad** Format **(*.dwg)** &
     3) **Facilities layers** (Multiple Entities) in **Intergraph** format **(*.dgn)**
     4) **UBD*/Drainage Raster Files** in **TGA ,PCX** and **Tiff** formats
2) Given a derived (x,y,z) in space from **Lot:Section:DP** of a parcel, display all **relevant layers** at a scale of **1:2000** with the **Drainage Plan** and the **unique_id's** of all LandParcels within **the current viewport**(displayed at each parcel's paracentroid with the derived **Land Title/Valuer General reports**
3) **Fax** to 02-438-4788

The mosaix macro environment is best illustrated with examples of the richness of functions. Illustrated below is a quite basic requirement of a G/LIS system:

The macro is very skeletal in nature but it does provide a solution within approximately **60 lines of code** from scratch - it should be noted that the macro is **command interpreted** and therefore does **not have to be compiled**

atable test; **#Define a Global Application Table**

group ~test.autocad;**# define some graphic groups for each format type**

group ,~test.intergraph,~test.landmaster

database ~test.valgen,-name='valgen',-pro='ingres' ; **# Open the Databases**

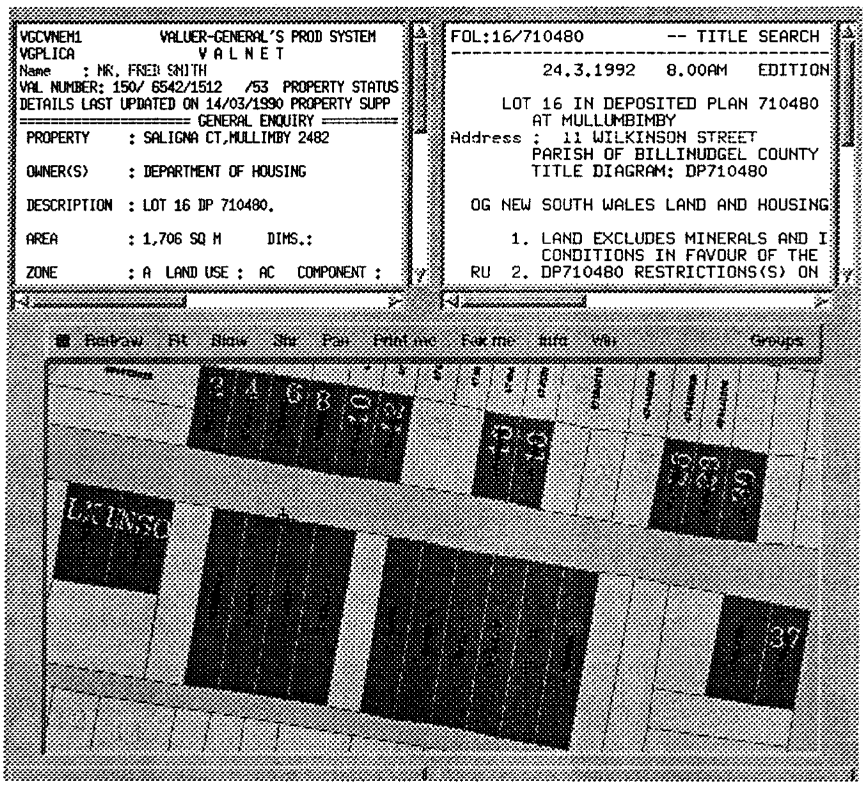
database ~test.landtitle,-name='landtitle',-pro='oracle'

database ~test.mapman,-name='mapman',-pro='informix';**# the Map Manager**

open landmaster, -s='landmaster_server';**# start the mosaix server processes**

open autocad, -s='autocad_server'

open intergraph, -s='intergraph_server'



window ~test.overview=wgraphic; **# create a graphic window for the overview**

gc=p0 ;**# position graphic cursor at (0,0,0)**

~test.ubdscale=6.0476 ;**# say 6.0476 metres per pixel scanned at 100dpi**

~test.mapt65=(151.124,-33.522) ;**# Origin of Bottoml.h. corner of UBDmap65.**

 # sydney.clo - Example cloth definition file- Notice how rasters are 'instanced'

level=1, scale=~test.ubdscale, view=plan

file = $RASTERS/xras/map650.ras, p=~test.mapt65, title='Page 65'

file = $RASTERS/xras/map550.ras, p=~test.mapt55, title='Page 55'

~test.overview.cloth='$RASTERS/sydney.clo' ; # **add the backcloth to the overview window**

wscale ~test.overview, -sc=2000,-p=(151.2,-33.5032);**#Prepare window @1:2000**

point window_extents[]=~test.overview.corners

 # create a 3d vector array (window_extents[]) with the long,lats of the corners of the newly created window

 # Enquire from the Informix DB, which files should be gathered for analysis -

 # let's say the generated SQL (see later example on db blocks) command sends us ('fetches') back a list of the full pathnames of object files that are 'fenced' by the window geometry into a string array ~test.files[]

2 # Intergraph *.dgn files

3 # LANDmaster *.mod files &

6 # Autocad *.dwg files

i=0

loop 1,~test.files.alength {

 # i.e. for the no of objects in the array back from MapManagement system

i++; **# this is just the same as saying i=i+1 for the old fortran hacks**

these=null ;**# initialise the incoming graphic group**

switch fext(~test.files[i]){ ;**# i.e. extract the file extension**

\#

 # each server will create named,datestamped entities in mosaix memory

 # from the binary format i.e. no translation to ascii required

case dgn: { write intergraph,~test.files[i]; ~test.intergraph += these}

case dwg: { write autocad,~test.files[i];~test.autocad += these}

case mod: { write landmaster,~test.files[i] ;~test.landmaster += these}

case other: { tell 'I don"t know what to do with [^(~test.files[i])]'}

}

}

#now place the unique_id at the paracentroid of all the LandParcels clashing with the current window

group poly_in_window=clump_w(~test.overview);**# find all the polys in window**

gloop poly_in_window { gtext this.name ,-c=2,-p=this.centroid};**# now, for each polygon in the resultant group, find the centroid and place text at this point with a height of 2m ground units**

wdisplay $RASTERS/drainage_1234.ras; **# display the drainage diag. (raster)**

walpha $REPORTS/landtitle_1234.rep;**# display the LandTitle Report (alpha)**

walpha $REPORTS/valgen_1234.rep;**# display the ValGen Report (alpha)**

 # dump the view to Postcript and filter through OZfax software to the modem

dump -iw=current_window,-uof='pl2ps |sendfax -shptl -t024384788'

1.1 mosaiX Language Features

C-Like blockstructured language comprising commands, control statments, definitions and variable assignments

Consistent command syntax with positional and optional parameters

Event driven input from on-screen menus,windows, keyboard and/or external processes

System defined data structures and variables providing consistent and full access to modelling, Unix, window manager, external process and database environments

1.2 Object/Identifiers

Global,local or application scopes

All objects are timed stamped

Predefined system structure types

channel	communication with external files/processes
window	communication with window manager
atable	application object definitions and environment
database	communication with external databases
cursor	communication with on screen input

1.3 Graphic Object types

entity	3-D graphic vector primitives (lines,ellipses,solids,polygons etc)
group	classification for primitives
filter	selection criteria for primitives (e.g all arcs digitised >time t1 & <t2)

1.4 Symbology structure types:

colour,linestyle,font,textstyle,markerstyle,tile,hatchstyle,fillstyle

1.5 User defined data structures:

structure	combination of non-graphic objects
dbstructure	transparent mapping to external database records
gstructure	structure associated with a graphic object
compound	assembly of graphic objects

1.6 Object Orientation concepts:

inheritance	structures inherit members and values from parent class
encapsulation	via structure member protection
polymorphism	via definable structure procedure member

class constructor and destructor procedures

1.7 Windows: Compliance with X-11 Toolkit Release 3/4/5

Compliance with Motif

Available window types:

graphic	display of Vector and Raster files
alpha	display of text files,help (in concurrent foreign languages) etc

forms   status,database communication etc

menu   alpha,icon,raster,toggle buttons etc

message   user information

display   image viewer

panel   widgets,dials ,sliders etc

ability for external processes to address window contents

1.8 Channels

Communication with external text data files. Formatted read/write operations.

Communication with external processes for sending/receiving data and commands

Ability to recieve commands from external processes on a polling/interrupt basis

Communication with external data servers for fast data import/export via supplied C libraries.

Definable event procedures for channel data selection

Read time out facilities

1.9 External Database

Seamless links to external proprietary databases (Informix,Oracle,Ingres)

Ability to send commands to external database and to read data direct into language variables

Ability to send commands to external database on a transaction/commit basis

Definable data structures (dbstructures) providing a transparent link to single or multiple records within the external database. Updating of data at one end of the link is reflected immediately at the other, or can be deffered

1.10 Symbology - All symbologies are named

1.11 File Management -Full access to unix file system via internal commands, unix shell commands or interactive file selection menu box

1.12 Licensing

Ability for developers to license(floating) their own products.

1.13 Database Event Capabilities

Products such as Ingres provide eventing mechanisms through extensions to SQL. The event is a named object/resource which is stored and managed within the database server. Applications such as windows 4GL and mosaiX can then elect to listen to an event or multi events. Applications can also elect to generate events from within their application code. When the application generates an event it can, optionally, pass a text string with the event. All applications that have asked to listen to the particular event will then be notified of the event and passed the event text.

We see this feature as one of the most exciting and significant 'glueing' facilities provided to GIS systems and databases (e.g. Real-time network display and analysis, Intelligent notification handling)

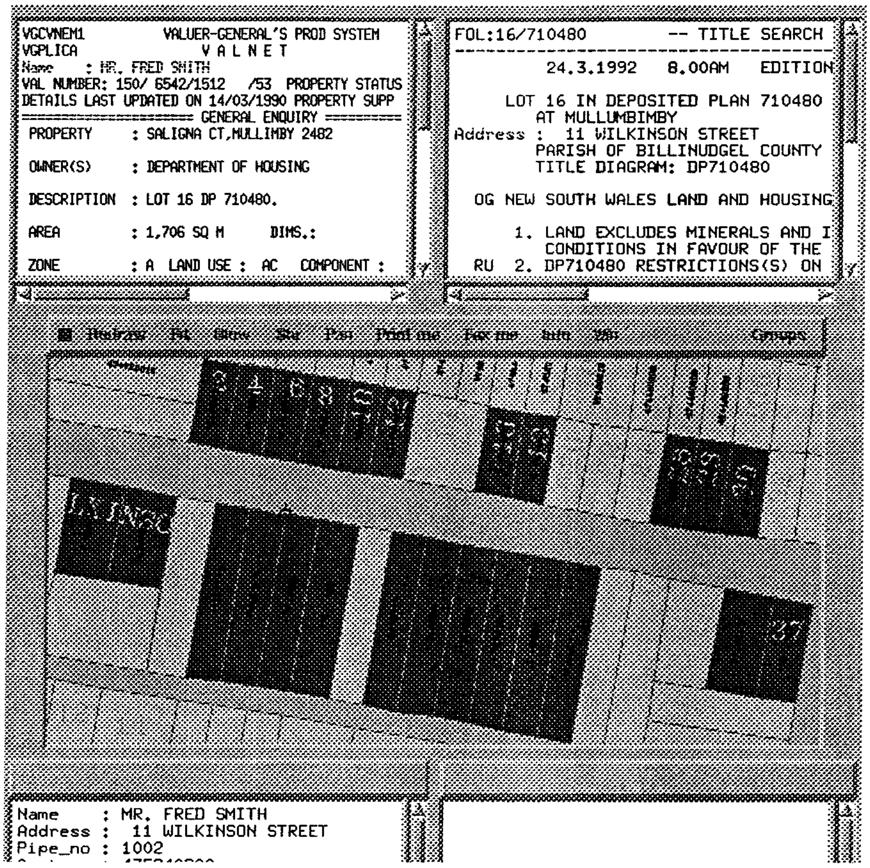
2.0 Application Library Interface

The mosaiX API consists of a set of 'C' language callable libraries which enables the developer to converse with a mosaiX session. The external program is typically initiated by the mosaiX application code which establishes a bi-directional communications channel to the process and the process is then passed the address of a shared memory segment. The bi-directional communications channel is used to synchronise the process with mosaix whilst the shared memory segment is used for the transfer of data and messages. No knowledge of the operations of shared memory interfaces is required by the developer and only a single mosaiX API call is required to initiate the shared memory interface. (Refer to macro example section 3.1) Having established a connection with mosaiX the external process has full access to the following functionality.

- Query any data value active within the mosaiX application.
- Set any data value.
- Execute any mosaiX command or function.
- Create,modify or delete any graphic entity.
- Respond to any commands sent from the application.(e.g. SCADA)
- Trigger an event for mosaiX to respond to.(e.g. from Windows 4GL)

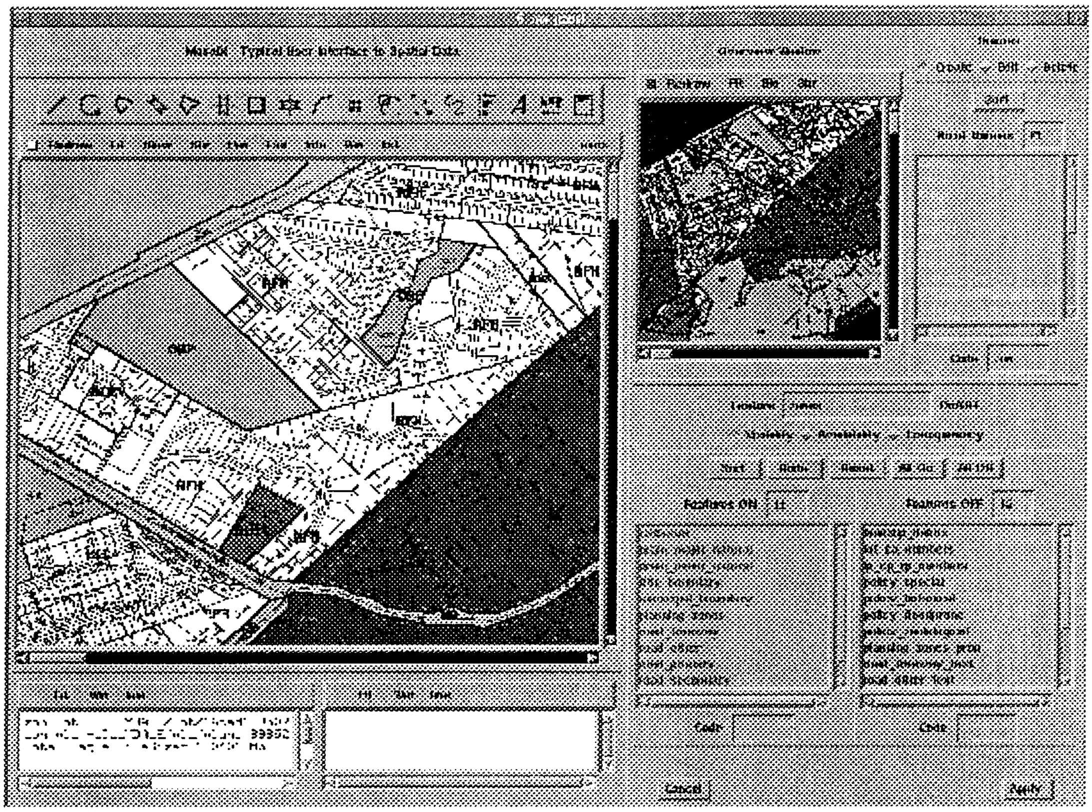
3.0 - Examples in Action

- 3.1) Incorporating SQL via DB Blocks.
- 3.2) Dynamic Translation of proprietary formats.
- 3.3) Object clashing and reporting with Buffer zones.
- 3.4) Schematic/Real World duality displaying 'raised' external events
- 3.5) Dynamic adjustment of object position from external source
- 3.6) Pulldown menu code example
- 3.7) More Complex Panel Construction
- 3.8) Using Charting objects
- 3.9) Call-Backs (Handlers) for dials to extrapolate pollution plumes generated by 3rd Party processes
- 3.10) 3-D Display of 3rd Party data from Digital Terrain Model Software
- 3.11) Displaying Aspatial Data (Business Application)



EXAMPLE 3.1: SAMPLE INTEGRATED SQL WITHIN MOSAIX MACRO CODE

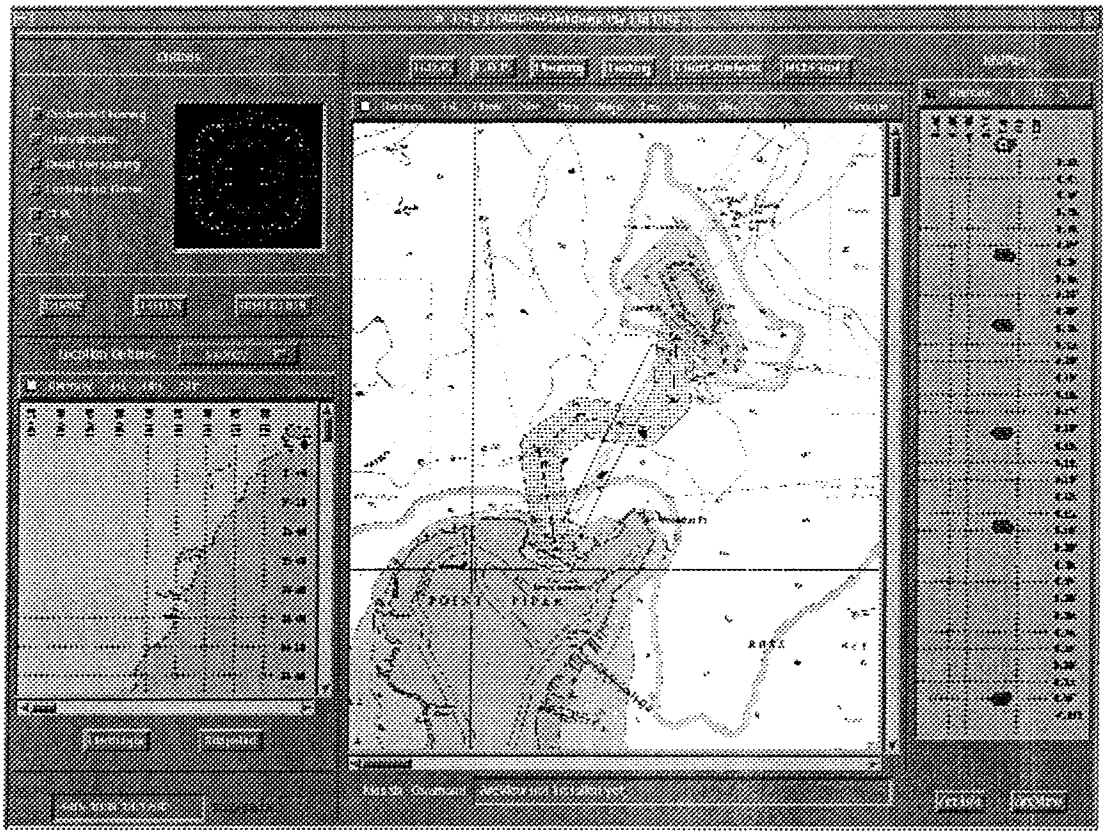
```
database ~agl.test,-name='agln',-pro='informix'
ask 'Enter SQL query for "SELECT poly from account where"...',sqlline ,-l,-d='street_nam-
e="WILKINSON"'
 if [ sqlline != '' ]{
db ~agl.test {
        * select u_poly from account where ^(sqlline)
        * fetch u_poly
        while [found] {
        if [exist(^(u_poly))] {
        ~agl.sqlresult = ~agl.sqlresult + ^(u_poly)
        }
        * fetch u_poly
 }
gloop ~agl.sqlresult {modify this .colour=blue}
point pts[2]=extents( ~agl.sqlresult)
box -p1=pts[1],-p2=pts[2];plan}
```



EXAMPLE 3.2: DYNAMIC TRANSLATION OF PROPRIETARY FORMATS

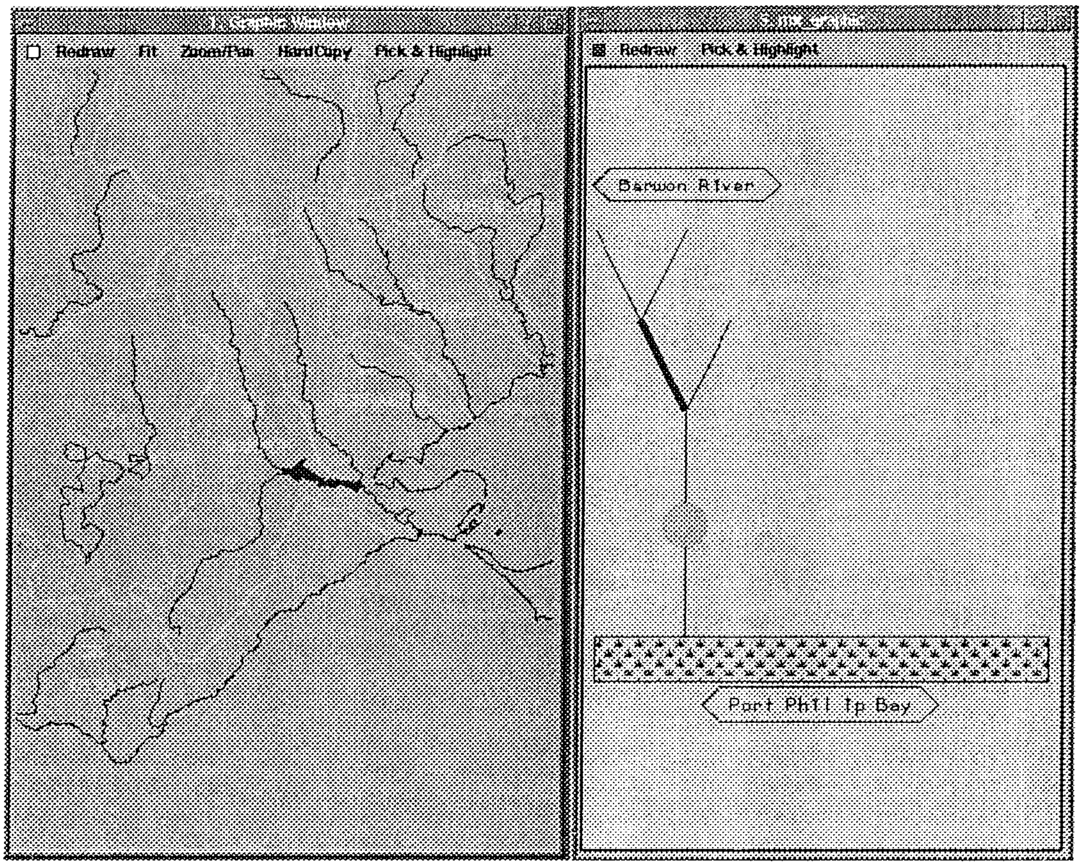
... We are supplied with an array of files (~test.files) to be displayed between the bottom left hand corner and top right hand corner of the current window - they just happen to be in two kinds of prorietary formats (AUTOCAD and LANDmaster)

```
 open landmaster, -s='landmaster_server';# start the mosaix server processes
open autocad, -s='autocad_server'; # -s denotes a shared memory process
i=0
loop 1,~test.files.alength {
# i.e. for the no of objects in the array back from MapManagement system
i++; # this is just the same as saying i=i+1 for the old fortran hacks
these=null ;# initialise the incoming graphic group
 switch fext(~test.files[i]){ ;# i.e. extract the file extension
# each server will create named,date-stamped entities in mosaix memory
# from the binary format i.e. no translation to ascii required
 case dwg: { write autocad,~test.files[i];~test.autocad += these}
 case mod: { write landmaster,~test.files[i] ;~test.landmaster += these}
 case other: { tell 'I don"t know what to do with [^(~test.files[i])]'}
       }
}
```



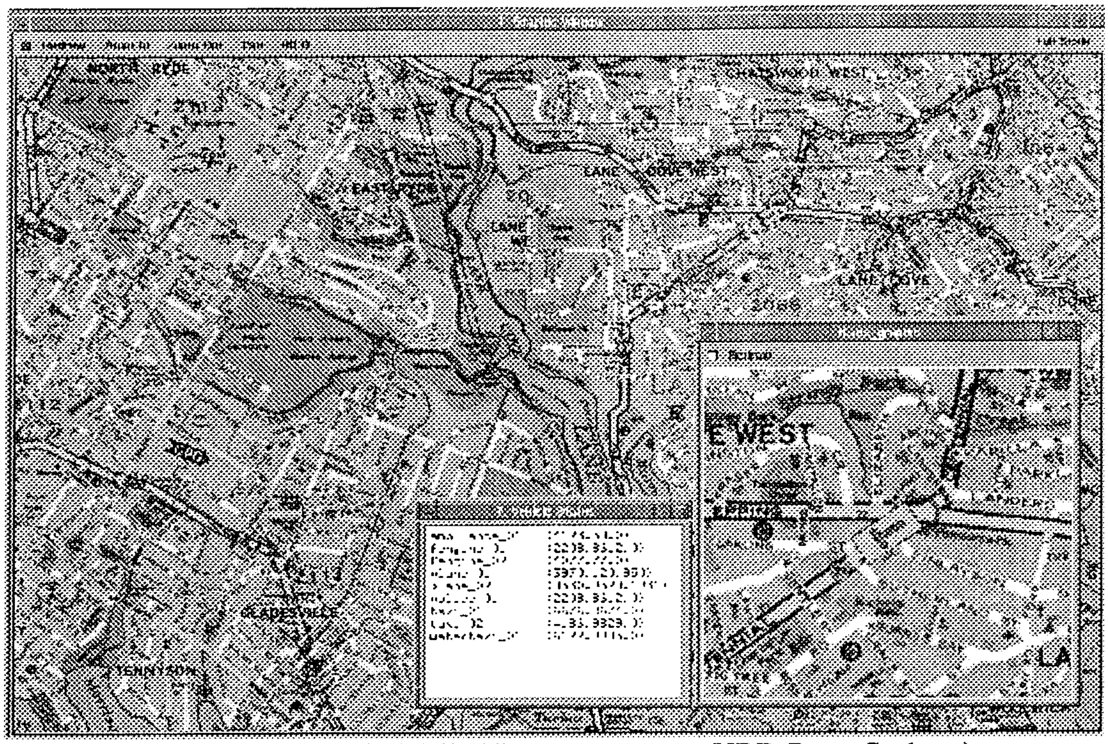
EXAMPLE 3.3: OBJECT CLASHING & REPORTING WITHIN BUFFER ZONES

```
numeric bufwidth ;          # declare a numeric variable bufwidth
group ~test.results;        # declare a group of pointers of the objects clashing
                            within the newly created buffer_zone
group buffer;               # declare a group called buffer
ask 'Enter Buffer Width',bufwidth,-d=150; # set a default width of 150 metres
zone bufwidth,-g=buffer,-col=blue,-fil=grey25 { pseries }
                            # Create a buffer zone with a series of interative points
~test.results = clash_p ( this.vertex )
                            # the last created entity is generically called 'this' -
                            therefore this.vertex is an arrray of the polygon making
                            up the buffer zone
                            # clash_p is a standard Mosaix Filter Function to return
                            all entities which clash with a polygon in 3-D
gloop ~test.results { tell this.name }
                            # for every entity in the group ~test.results, tell the user
                            it's name
```



EXAMPLE 3.4: SCHEMATIC/ REAL WORLD DUALITY with EVENT DRIVEN DATABASE

```
database water,-prop=ingres,-name,'supply'; # connect to the database
water.exec = { # Define dbevent callback
        args db=database,event=string,event_text=string
        switch event {
                case 'flood_warning' {
                        ...;tell event_text;...}
                case 'algae_warning' {
                        ...;tell event_text;...}
                other { tell ' Got a dbevent that is uncatered for.'}
}       }
db water {  *   register dbevent flood_warning ; # Register for dbevents
            *   register dbevent algae_warning
            *   register dbevent new_supply_order }
water.eventing = on ; # activate database eventing
# We may wish to register events from mosiax instead of from Ingres
db water {  *   remove dbevent new_supply_order;# now we are going to 'RAISE' an event
            ........;# some mosaix and SQL code which creates a new_supply_order
            *   raise dbevent new_supply_order '123456,some text' }
```

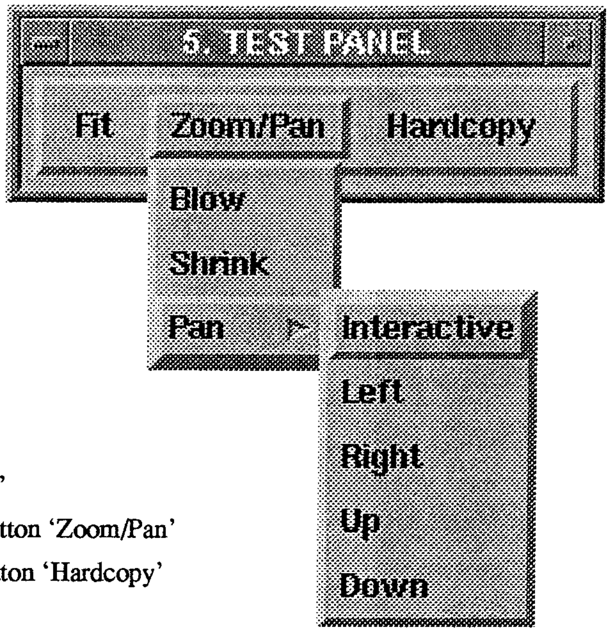


(Raster Backcloth Tiled Images courtesy UBD Press Sydney)

EXAMPLE 3.5: DYNAMIC ADJUSTMENT OF AN OBJECT (Vehicle) POSITION FROM AN EXTERNAL DATA SOURCE

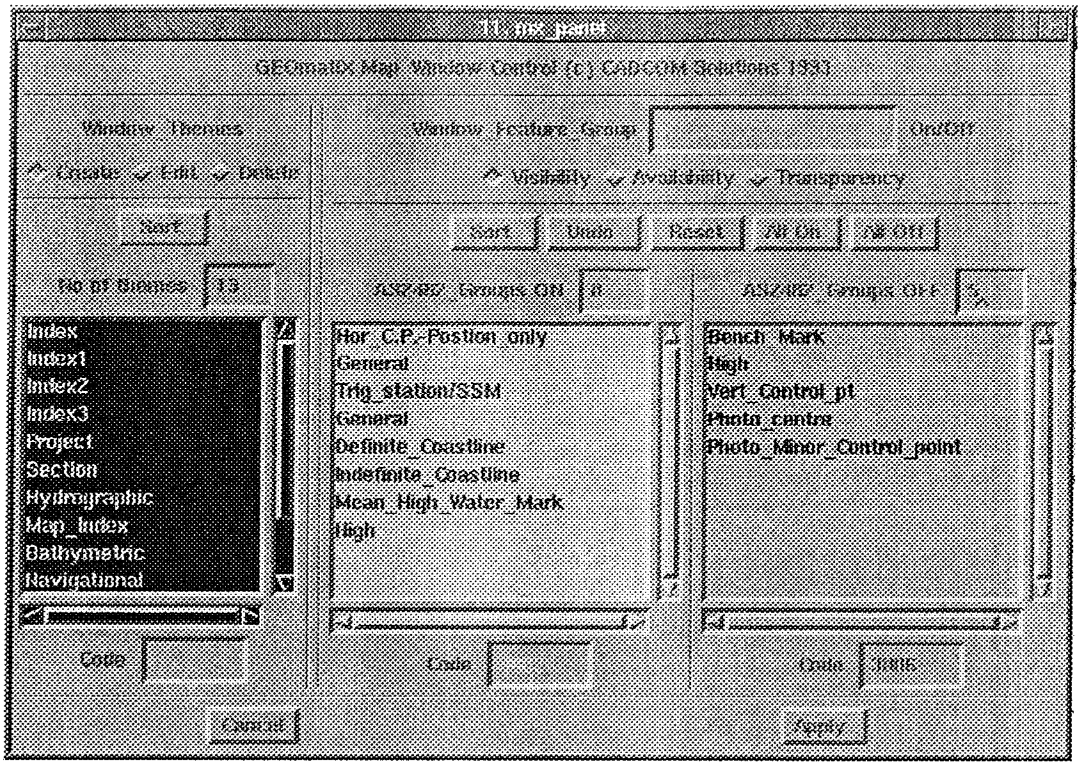
```
# FUNCTION: Define a vehicle reposition function
 function ~ubd.move_vehicle {
 args veh = vehicle, pos = point

# Set the position and angle of the graphics
        veh.inst.angle = ang(veh.origin,pos)
        modify veh { origin = pos }
# Set its position in monitor window
        listwin.lines[veh.info_row]%16:* = ps(pos)
        if [ veh.info_row = hilited ] {
#
# Pan the screen if "VEH" moves out of ~ubd.zoom_window
        if [zoom_window] {
                point pc[2] = zoom_window.corners
                if [ inbox(pos,pc[1],pc[2]) = 0 ] {
                window cw = current_window
                current_window = ~ubd.zoom_window
                wopen zoom_window
                wexpose zoom_window
                pan -p=pos,-sup
                zoom_window.cloth_mag = 2;current_window=cw }}}}
```




```
window test=wpanel {
        row -mb {
                fit: button 'Fit'
                zoom_pan: button 'Zoom/Pan'
                Hardcopy: button 'Hardcopy'
                }
        }
        test.title='TEST PANEL'
test.zoom_pan.exec =
        pulldown {
                1: 'Blow' { blow }
                2: 'Shrink' { shrink }
                3: 'Pan' {
                        pulldown {
                                1: 'Interactive' { pan }
                                2: 'Left' { pan -l=0.5 }
                                3: 'Right' { pan -r=0.5 }
                                4: 'Up' { pan -u=0.5 }
                                5: 'Down' { pan -d=0.5 }
                                }
                        }
                }
        }
```

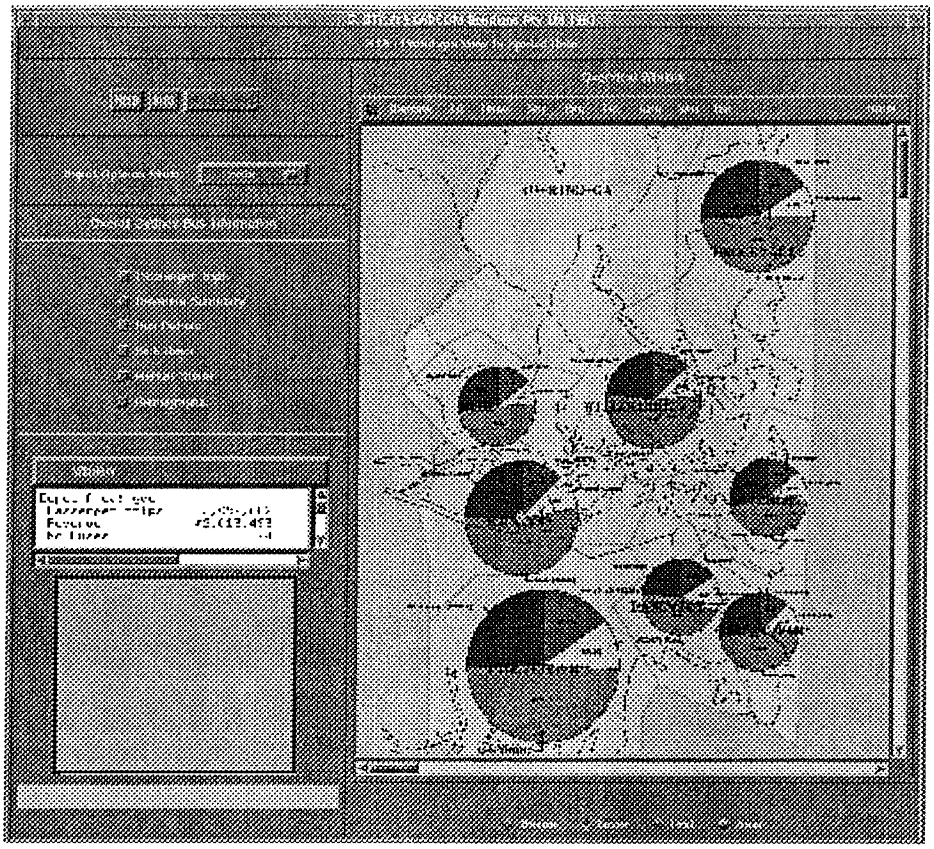
EXAMPLE 3.6: SIMPLE PULLDOWN MENU CODE



EXAMPLE 3.7 : MORE COMPLEX PANEL CONSTRUCTION

This macro code wil generate the first column of the above panel:

```
column {
    theme_title: label 'Window_Themes'
    row -eq { theme_status: radio_box <'Create','Edit','Delete'>,-h}
    separator
    row { theme_sort: button ' Sort '}
    gap 10
    row {
        theme_num_title: label 'No of themes'
        theme_num_box: text ns(~sis.theme_cod.alength),-out,-c=4
        }
    theme_list: list ~sis.theme_desc ,-bc=blue,r=10,-w=150,tc=~sis.theme_col
    row {
        theme_code_onl: label 'Code'
        theme_code_on: text '',-out,-c=7
    }
}
separator;separator
...;# now setup the 'call-back' when theme_list is double-clicked i.e. add a code to 'code_on'
~test.panel.theme_list.exec {
~test.panel.theme_code_on.text = ~test.panel.theme_code[_~test.panel.theme_list.value]}
```

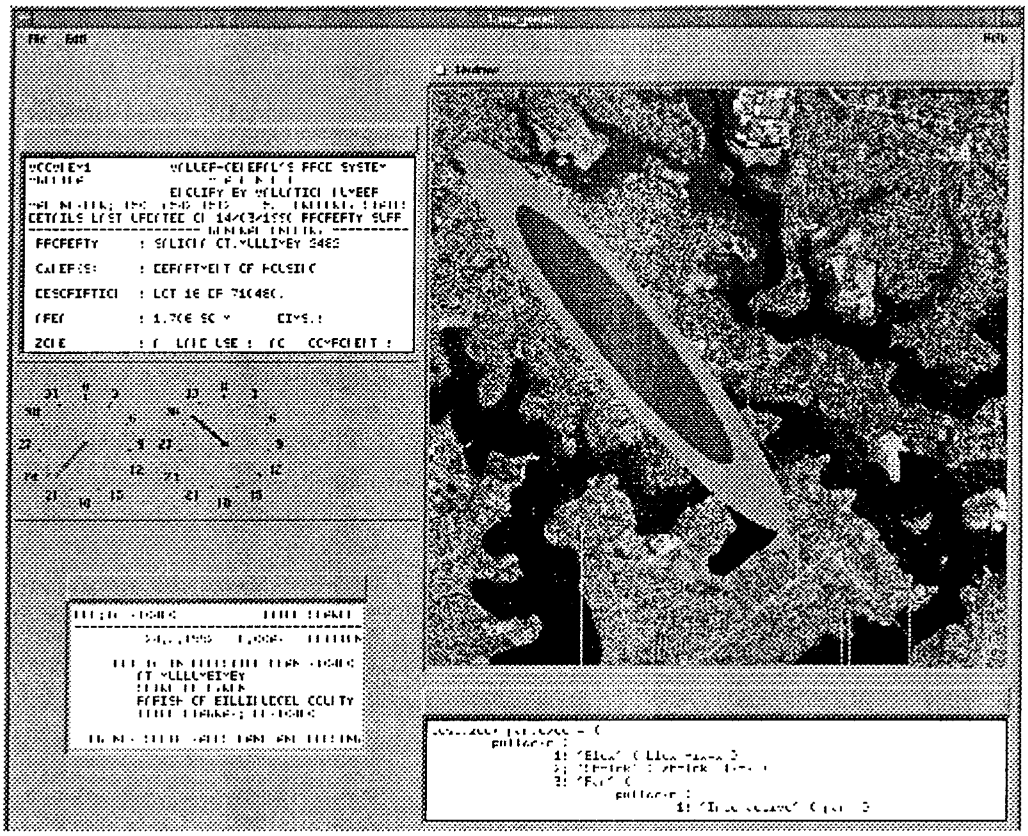


EXAMPLE 3.8: USING CHARTING OBJECTS

For each polygonal area (~test.area[]) we have data of
total number of responses (e.g. ~test.waverley=36000) for
four age groups (e.g. ~test.ages[1]=11-14 yr old,~test.ages[2]=15-21 etc)

for area_no=1,~test.zones.alength {;# **for each area**
 # area_no must be converted to a name (e.g.area [1] = waverley)
string area_name=^(~test.area[area_no])
chart_origin=^(area_name).centroid;# **find the centroid of the polygon in question**
chart_radius=biggest_circle_possible*sqrt(total_per_area[area_no]/biggest)
 # make radius a proportion of the biggest radius to be displayed

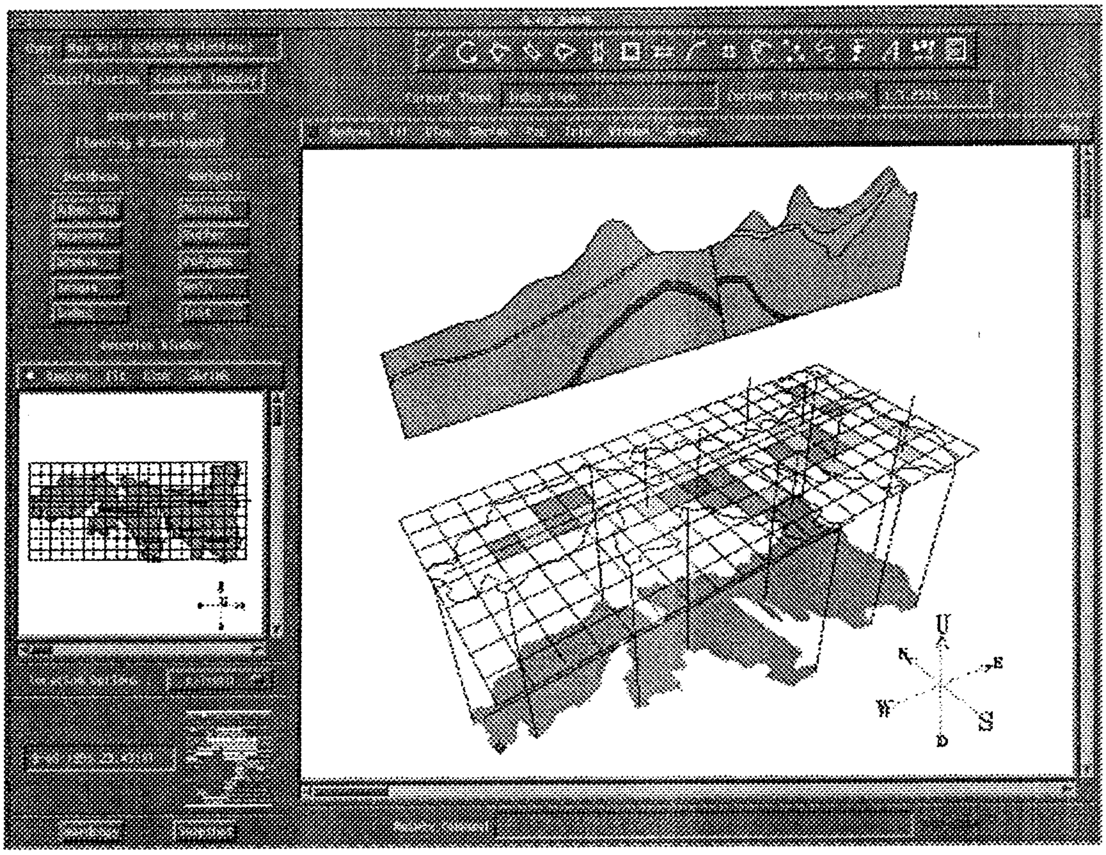
~graph.pie_chart(area_name,'response vs age',chart_origin,chart_radius);
 # create the piechart of [area_name] with [response vs age] at the polygon _centroid with a nominated radius
}



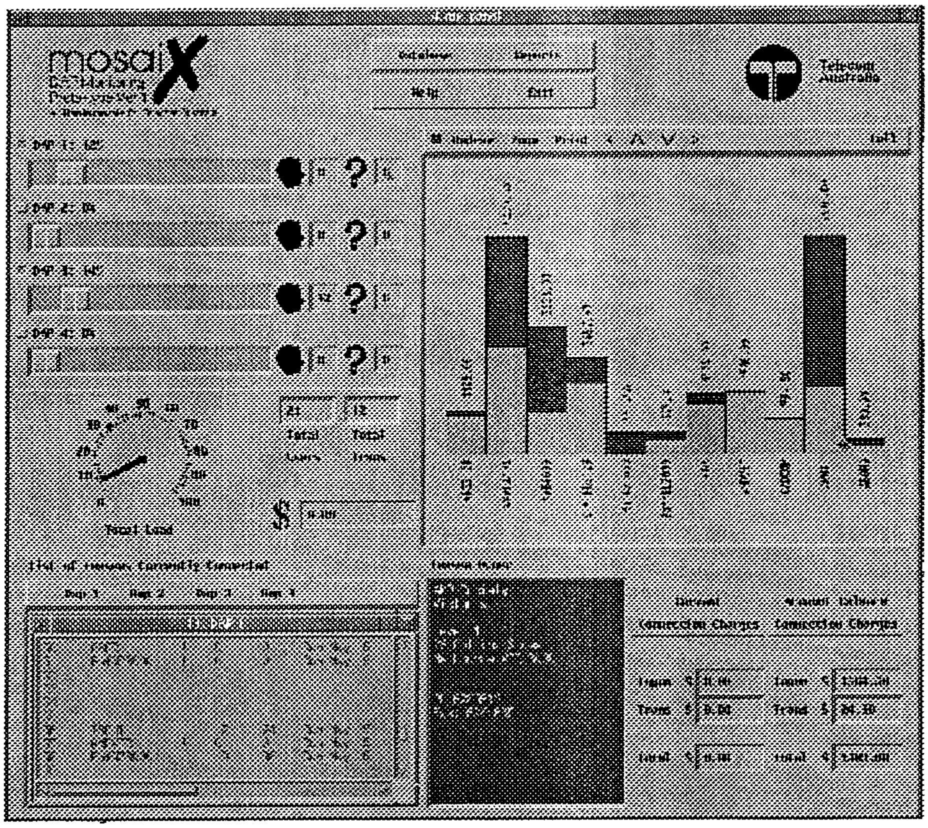
EXAMPLE 3.9 : CALL_BACKS (HANDLERS) FOR DIALS TO EXTRAPOLATE POLLUTION PLUMES GENERATED BY 3rd PARTY PROCESSES

```
~test.pollution.wind_speed.drag={
 args w=window
 tell <w.wind_speed.value,~test.pollution.wind_direction.value>
... call external process to calculate pollution plume
... mosaix function to dynamically translate the output to graphics as in example
}
~test.pollution.wind_direction.drag={
 args w=window
tell <~test.pollution.wind_speed.value,w.wind_direction.value>
..... call external process to calculate pollution plume
..... mosaix function to dynamically translate the output to graphics as in example
}
```

(Raster Backcloth courtesy SPOT Imaging, Sydney)



EXAMPLE 3.10 : 3-D DISPLAY of DIGITAL TERRAIN MODEL



EXAMPLE 3.11 : DISPLAY ASPATIAL DATA (BUSINESS APPLICATION)

CONCLUDING REMARKS

mosaiX does not have to carry the 'baggage' of an historical client base. It is 90's software. It attempts to provide a truly integrated system and solve some of the existing barriers of data conversion,ongoing data maintenance, application software delivery and data distribution to users.

ACKNOWLEDGEMENTS and REFERENCES

M.W.McEwan (1993) -Various white papers on mosaiX Technology

Murnane,B. (1992); **INTEGRATING SPATIAL SYSTEMS**; PROCEEDINGS Aurisa 92.

CARTOGRAPHIC 3-D FLY AROUND TERRAIN MODEL MOUNT CANOBALAS, NEW SOUTH WALES

Mr John P. Hindle
Agricultural Environment Officer
NSW Agriculture
PO Box 393,Nowra 2541
Australia

ABSTRACT

In this study, a digital elevation model (DEM) and a land cover image were used to create a cartographic 3-D fly around terrain model of Mount Canobalas, New South Wales. The DEM was generated in an analytical plotter (Wild Aviolyt BC2), from Spot HRV-7 stereo multispectral colour positive transparencies, while the land cover image and 3-D images were created with image analysis and spatial information system (SIS) software from Spot HRV-7 multispectral digital data.

A unique feature of this research is the use of automatic computer processing capabilities to create 3-D images and the automatic display of 3-D images on the computer screen for recording by a high frequency video camera. The 3-D images were created in a SIS by draping the land cover images over the 3-D elevation model to produce 3-D images of land cover features. A series of these 3-D images were created for each two (2) degrees of compass direction. These images were displayed on the computer screen for recording with a high frequency video camera. Editing of the one hundred and eighty (180) recorded 3-D images provided the visual rotation of the 3-D terrain model of Mount Canobolas, New South Wales.

This research demonstrates the state of the art for rotating 3-D terrain models with complex land cover patterns. 3-D objects or models with uniform coloured lines or surfaces can be rotated in real time using Silicon Graphics work stations. However, further research is necessary to rotate 3-D terrain models with complex land cover patterns in real time.

INTRODUCTION

The animation process provides the cartographer with a means of visually displaying and exploring real world phenomena, particularly where the date contains spatial and/or temporal interactions, and recognisable sequences of change within environmental and natural resource systems.

This animation process has wide application in extracting meaningful data from real world phenomena, by providing the cartographer with the ability to communicate information to a wide audience in a realistic manner. Here, the animation process and 3-D displays can be used to create movement, replacing the traditional static 2-D maps.

3-D modelling techniques are currently used by meteorologists, medical researches and physical scientists to formulate hypothesis about recognisable sequences of change and variability within environmental and natural resource systems.

This research uses the state of the art technique for rotating 3-D terrain models with complex land cover patterns. Simple objects with uniform coloured surfaces can be rotated in real time using a Silicon Graphics work station. However, to rotate 3-D terrain models with multi-coloured surfaces in real time, further research is necessary particularly where remote sensed data is draped across the wire frame 3-D model.

The aim of this project is to prepare a cartographic 3-D fly around terrain model of Mount Canobalas, New South Wales, using digital elevation model (DEM) and land cover data obtained from Spot HRV-7 multispectral colour positive transparency images and digital data.

LITERATURE REVIEW

At present spatial data collected from environmental processes and natural resources are generally displayed as static 2-D maps. With the development of cartographic modelling techniques, animation offers an alternative method of viewing sequences of change and formulating hypothesis about the spatial distribution and interactions within natural phenomenon (Weber, 1991,p.2).

Here, huge quantities of data collected from remote sensing devices and analysed by image classification techniques and/or spatial information systems (SIS), (Marshall, Kempf, Dyer and Yen, 1990,p.89) can be interactively explored by rotation of the 3-D model (Hibbard, Uccellini, Santek and Brill, 1989,p.394).

These visual displays can be used to communicate 3-D information (Hibbard, 1986,p.1362) with their success depending on the viewing environment (Carter, 1988,p.379) particularly the use of colour (Murch, 1985,p.1362). Well chosen colours will increase the attractiveness of a map or 3-D model and the effectiveness of communicating spatial information (Brewer, 1989,p.269). This particularly applies to highlighting the detail (Olson, 1981,p.276) of external and internal surfaces of objects (Rheingans and Tebbs, 1991,p.145).

Perhaps, one of the major features of automation is the provision of viewer interaction with maps and statistical diagrams as spatial and temporal sequences (Monmonier, 1990,p.30), particularly where the display sequences can be automated by writing sequence files called graphic scripts (Monmonier, 1989,p.381). This provides an opportunity to analyse interactions with resolution of specific problems (Moellering, 1980,p.12) or as an analytical tool to formulate hypothesis about sequences of change within natural phenomena. This is achieved by analysis of interactions between variables (Grotjahn and Chervin, 1984,p.1201) within natural processes (Hibbard, et al, 1989,p.1394) .

Study Site

The study site is located south west of Orange, on the Central Tablelands of New South Wales. It includes the town of Orange, extensive areas of deciduous orchards, pine plantations and the foothills of Mount Canobolas. It has a diverse range of landform, vegetation and land cover features.

Data Set

A pair of stereo cloud free Spot HRV-7 multispectral colour positive transparency imagery (1-B level) and Computer Compatible Tapes (CCT's) were acquired for the study area on 20/3/88 and 25/4/88 with look directions of R 27°31' and L 17°6', elevation of 43°6', 37°4' and azimuth of 51°6' and 31°9' respectively. The CCT image data (20/3/88) was rectified to remove the stereo affect, allowing the image data to be registered to a map base prior to the classification of land cover features.

METHODOLOGY

A DEM was produced in a Wild Aviolyt Analytical Plotter (BC2) at the University of New South Wales from a pair of Spot HRV-7 multispectral colour positive transparency images. Digital data for one of the Spot HRV-7 multispectral colour positive transparency images (20/3/88) was classified by the Maximum Likelihood Classifier to produce a land classification image. This image was draped over a wire frame terrain model to create a 3-D terrain model of Mount Canobolas and its foothills.

The DEM data was processed by Arc Info® software, while the land cover image was produced by the Erdas® image analysis software. The 3-D terrain images for each two (2) degrees of compass rotation were processed by Arc Info® software and displayed on the Sun Sparc Station at the University of New South Wales.

Each of the single 3-D terrain model images were captured by a high frequency video camera and recorded on a video tape. Editing of the video tape provided six (6) frames per 3-D terrain model image. This created a fly around 3-D terrain model of Mount Canobolas when viewed on a video screen.

RESULTS

The methodology successfully created a cartographic 3-D fly around model of Mount Canobolas, New South Wales. This 3-D model also displays and conveys information highlighting the spatial location of land cover patterns in relation to terrain features.

DISCUSSION

The classification of Spot HRV-7 multispectral image data acquired on the 20/3/88 into land cover classes produced slight misclassification of orchards, eucalyptus forests and urban areas containing ornamental trees and lawns. This was primarily due to similar reflectance count values being recorded for the infrared band.

This misclassification can also be partly explained by the large dynamic range within the Spot Image data compared with Landsat MSS image data.

In this research cartography was used promote the use of temporal and categorical variables where continuous variables occur in the real world. This was achieved by placing emphasis on the spatial characteristics of mapped phenomena as the primary factor in symbol selection. Here, the producer uses geometric progression and the choice of colour to determine the cartographer's subjective impression of the data.

Colour and variation of hue can quickly communicate information to map viewers. Here, care is needed in selecting colour as the perception of colour can vary as a result of variation of output devices, while the choice of colour can alter time and intensity perception.

In many cases, symbolic visualisation strategies are influenced by viewing time, position and viewer expectations.

CONCLUSION

A 3-D fly around model of Mount Canobolas has been successfully created to;

1. Demonstrate various production techniques
2. Highlight terrain features and orchard areas on the foothills of Mount Canobolas.

The production of intelligible and unambiguous 3-D terrain models and maps for use by individuals with a minimum of training in the extraction of meaningful data has been dismissed as unattainable. It has long been the role of cartographic research to defy this maxim. The most satisfactory 3-D terrain models and maps in communicating information are not the most realistic, but those which distort reality. This can be achieved by eliminating unnecessary information and exaggeration of useful data. The use of described colour in combination with a minimalistic map composition have been successful in presenting information to a maximum number of trained and untrained viewers.

Images produced from scientific data should not be viewed as the final analysis or as a methodology for displaying data, but as a communication technique for;

1. Explanation of data created from mathematical models
2. Examining huge quantities of data from laboratory experiments or remote sensing devices.

This ability provides scientists with an effective communication medium to increase their understanding and insight into spatiotemporal phenomena with better management of the environment and natural resources.

REFERENCES

Brewer, C.A. (1989) "The Development of Process-Printed Munsell Charts for Selecting Map Colon".
The American Cartographer, 16, 4, 269-278.

Carter, J.R. (1988) "The Map Viewing Environment: A Significant Factor in Cartographic Design".
The American Cartographer, 15, 4, Oct 379-385.

Grotjahn, R & Chervin, R.M. (1984) "Animated Graphics in Meteorological Research & Presentation"
Bulletin American Meteorological Society. Vol 65, No.11 1201-1208.

Hibbard, W.L. (1986). "Computer Generated Imagery for 4-D Meteorological Data"
Bulletin American Meteorological Society, Vol 67, 11, 1362-1369.

Hibbard, W.L. Uccellini, L., Santek, D. & Brill, K. (1989) "Application of the 4-D Mcldas to a Model Diagnostic Study of the Presidents' Cyclone".
Bulletin American Meteorological Society, Vol 70, 11, 1394-1403.

Marshall, R. Kempf, J. Dyer, S & Yen C.C. (1990) "Visualgation Methods and Simulation Steering for a 3-D Turbulence Model of Lake Erie".
Computer Graphics 24, 2, March 89-96.

Moellering, H. (1980,B) "The Real-Time Animation of Three-Dimensional Maps".
The American Cartographer, Vol 7, No.1, 67-75.

Monmonier, M, (1989) "Graphic Scripts for the Sequenced Visualisation of Geographic Data"
Proceedings GIS/LIS 89 Nov 26-30, 381-389.

Monmonier, M, (1990) "Strategies for the Visualisation of Geographic Time-Series Data".
Cartographica Vol 27, No.1 30-45.

Murch, G.M. (1985) "Colour Graphics-Blessing or Ballyhoo"
Computer Graphics Forum, 4, 127-135.

Olson, J.M., (1981) "Spectrally Encoded Two-Variable Maps"
Annuls of Association of American Geographics 259-276.

Rheingans, P & Tebbs B (1990) "A Tool for Dynamic Explanations of Colour Mappings".
Computer Graphics, 24, 2, March 145-146.

Weber, C.R. (1991) "A Cartographic Animation of Average Yearly Surface Temperatures for the 48 Contiguous United States: 1897-1986.
National Centre for Geographic Information and Analysis. Report 91-3.

OIL SPILL MANAGEMENT USING GIS

Gary A. Jeffress
Director of Geographic Information Science
The Conrad Blucher Institute for Surveying and Science
Corpus Christi State University
A Campus of The Texas A&M University System
6300 Ocean Drive,
Corpus Christi, Texas 78412
USA
Phone: (512) 994-2720
Fax: (512) 994-2715
EMail: gjeffres@falcon.ccsu.edu

ABSTRACT

Geographic Information Systems (GIS) have already proven their worth in the management of oil spills during spills from *Exxon Valdez*, 1989, *World Prodigy*, 1989, and the *American Trader*, 1990. In all of these cases GIS was used to monitor clean up efforts and to assess the extent of contamination caused by the spill. By using real-time environmental data in conjunction with environmental sensitivity index mapping, GIS will provide information that may reduce the risk of a spill and limit the extent of oil contamination in the event of a spill.

This paper describes a network of remote environmental sensing platforms along the Texas coast known as the Texas Coastal Ocean Observation Network (TCOON) TCOON is currently providing near-real time data useful for coastal navigation as well as providing critical data for oil spill trajectory models. These data are being entered in a GIS using ARC/INFO™ software at the Conrad Blucher Institute for Surveying and Science (CBI) at Corpus Christi State University.

INTRODUCTION

The Texas Coastal Ocean Observation Network (TCOON) presently comprises 44 data collection platforms (DCP) collecting a variety of environmental data in bays, estuaries, and offshore along the Texas Gulf coast. The establishment of this network was commenced in 1989 by the Conrad Blucher Institute for Surveying and Science (CBI) at Corpus Christi State University. The original network comprised three remote reading tide gauges along the shores of Corpus Christi Bay. The primary use of these gauges was to provide real time water level data for hurricane preparedness to the City of Corpus Christi. Data collected in near real-time from TCOON is now being integrated into a GIS under development by the Texas General Land Office's, Program for Oil Spill Prevention and Response.

From its roots as a real-time water level data system comprising the original three gauges, other users of water level data became interested in accessing data from the network. This led to cooperative agreements between the CBI and other parties to expand the network and the types of data collected. The focus of the network shifted from hurricane preparedness to water level measurement for use in the determination of littoral boundaries, and scientific research. Additionally, some DCP's collect data on wind speed, direction, water temperature, air temperature and barometric pressure. Among the current users and sponsors are the Texas General Land Office (TGLO), Texas Water Development Board (TWDB), the Environmental Protection Agency (EPA), and the National Oceanic and Atmospheric Administration's (NOAA) National Ocean Service (NOS). Due to the water level data being used for the determination of littoral boundaries, all gauges have been established to NOS standards. This allows water level data to be measured reliably and also addresses the legal concerns regarding the admissibility of the data for the determination of littoral boundaries (Jeffress, 1991).

Figure 1 shows the location of the gauges which presently comprise the TCOON.

Attention of the real-time data provided TCOON has now turned to its possible use in oil spill. This paper outlines the development of a pilot system to incorporate wind speed, wind direction, water temperature, air temperature and other environmental data into a real-time graphical display within ARC/INFO, for use at a potential oil spill site. An explanation of the way in which the data is collected in the Real Time Network (RTNET) is given. The data transfer method to the GIS data base and the implementation of the graphical user interface for ARC/INFO are briefly described. The role of GIS in the management of an oil spill response is then discussed and conclusions are drawn regarding the contribution the RTNET will have to the management of an oil spill (Garrett and Jeffress, 1992).

INSTRUMENTATION

The equipment which constitutes the TCOON are SUTRON 9000 and Vitel data loggers controlling Aquatrak acoustic sounders for the measurement of water level. The modular design of the SUTRON 9000 data logger allows the inclusion of additional sensors to measure other environmental and meteorological data. Using this feature, some of the SUTRON data loggers have been interfaced to wind speed and direction sensors, barometers and thermistors.

Currently, all stations within the network measure water level. Some DCP's house Hydrolab multi-parameter water quality sensing systems. These platforms provide information on water pH, salinity, dissolved oxygen and specific conductivity.

I refer the reader to Jeffress (1991) for a more detailed description of the TCOON. For the purposes of this paper, it will be sufficient to state that the data arrives at the CBI via at least one of three telecommunication channels. Most of the gauges in the TCOON transmit data via the Geostationary Operational Environmental Satellite (GOES) to the National Environmental Satellite, Data and Information Service (NESDIS), Wallops Island, Virginia. This mode is not real time but is the most reliable link CBI has to the gauges. Data from NESDIS is currently downloaded via telephone to the CBI automatically, every six hours. In addition to the GOES transmissions, many of the gauges also transmit data via telephone or packet radio in response to a 'poll' initiated automatically by computers at the CBI. This polling takes place every two hours. Five gauges which also have the ability to measure wind speed and direction, initiate their own data transfers every 8 or 15 minutes and these transmissions are received by the CBI. It is these five gauges which comprise the Real Time Network (RTNET) which provides the data for this pilot project. The data from all gauges is stored in digital form in a database at the CBI (Garrett and Jeffress, 1992).

Upgrading the Network

The TCOON is in the process of being upgraded. New software which is currently under development at the CBI will poll the majority of the gauges within the network every 15 minutes via packet radio. The number of gauges polled and the interval at which they are polled, will be able to be varied by the user. It will also be possible to poll individual gauges in an 'ad hoc' manner. This will allow a much more flexible approach to the data collection process. If a spill were to occur, gauges in the immediate vicinity of the spill could be identified and their status in the polling hierarchy increased. These changes will result in a significant improvement in the real time capabilities of the existing network. As stated above it is presently only the 5 beaconing gauges (i.e.. the gauges which initiate their own data transfers every 8/15 minutes), which can be considered 'real time'.

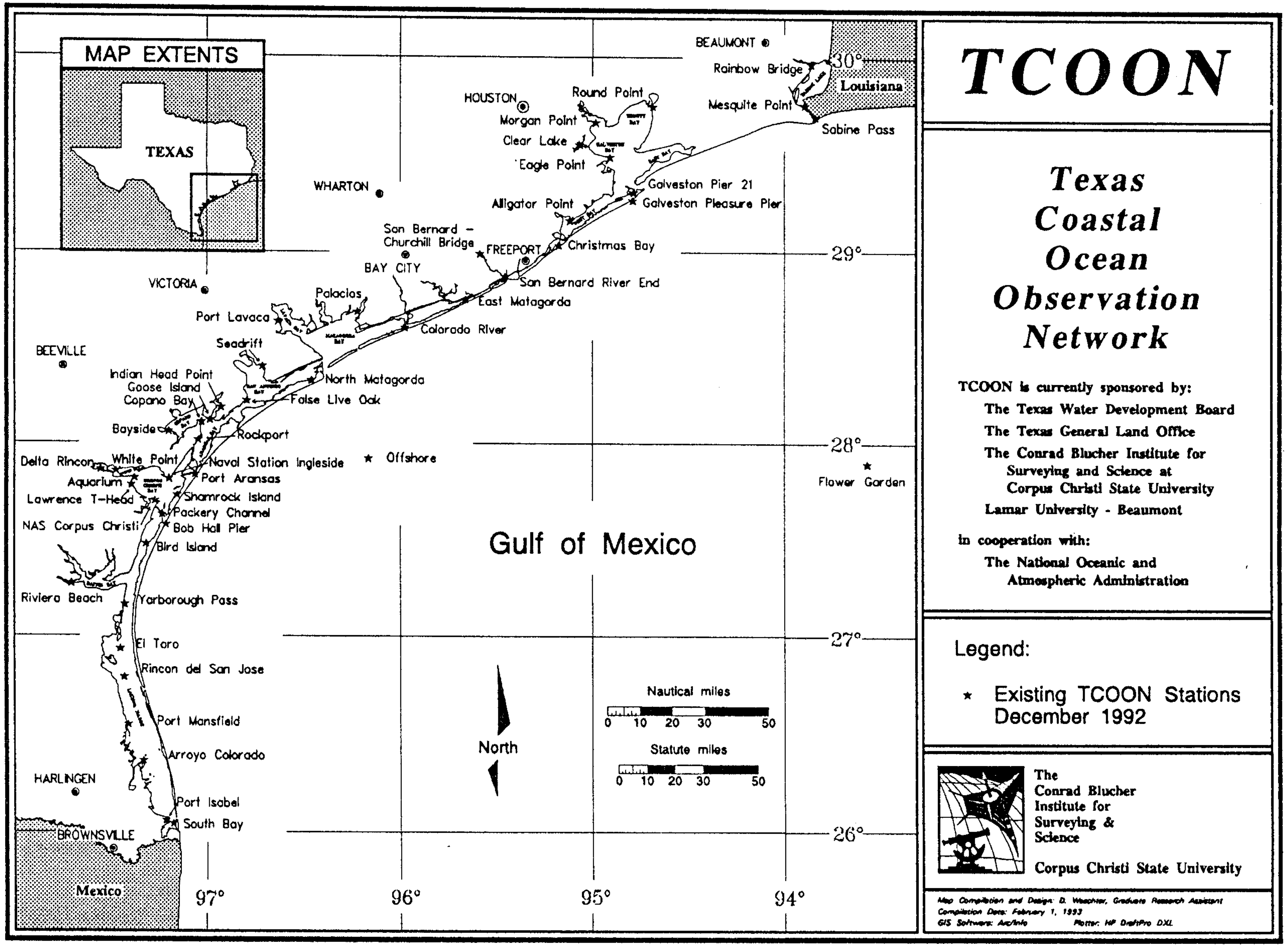

MAP EXTENTS
TEXAS
BEAUMONT
Rainbow Bridge
30°
Louisiana
Mesquite Point
Sabine Pass
HOUSTON
Round Point
Morgan Point
Clear Lake
Eagle Point
WHARTON
Galveston Pier 21
Galveston Pleasure Pier
Alligator Point
San Bernard -
Churchill Bridge
FREEPORT
Christmas Bay
BAY CITY
29°
VICTORIA
San Bernard River End
Palacios
East Matagorda
Port Lavaca
Colorado River
Seadrift
BEEVILLE
Indian Head Point
Goose Island
Copano Bay
North Matagorda
False Live Oak
Bayside
Rockport
28°
Delta Rincon
White Point
Naval Station Ingleside
Offshore
Flower Garden
Aquarium
Port Aransas
Lawrence T-Head
Shamrock Island
Packery Channel
NAS Corpus Christi
Bob Hall Pier
Bird Island
Gulf of Mexico
Riviera Beach
Yarborough Pass
27°
El Toro
Rincon del San Jose
Nautical miles
0 10 20 30 50
Port Mansfield
North
Arroyo Colorado
Statute miles
0 10 20 30 50
HARLINGEN
Port Isabel
South Bay
26°
BROWNSVILLE
Mexico
97°
96°
95°
94°
TCOON
Texas
Coastal
Ocean
Observation
Network
TCOON is currently sponsored by:
The Texas Water Development Board
The Texas General Land Office
The Conrad Blucher Institute for
Surveying and Science at
Corpus Christi State University
Lamar University - Beaumont
in cooperation with:
The National Oceanic and
Atmospheric Administration
Legend:
★ Existing TCOON Stations
December 1992
The
Conrad Blucher
Institute for
Surveying &
Science
Corpus Christi State University
Map Compilation and Design: D. Waechter, Graduate Research Assistant
Compilation Date: February 1, 1993
GIS Software: Arc/Info
Plotter: HP DraftPro DXL

Figure 1.

The addition of more radio frequencies would dramatically reduce the time required to poll the gauges in the network. A dual frequency system would mean that the CBI computers could poll and receive data from two gauges at the same time. If the network were to expand greatly beyond the currently projected 44 gauges, this may be necessary in order to reduce the polling cycle time through the network (Garrett and Jeffress, 1992).

REAL-TIME DATA AND OIL SPILL MODELING

There has been significant developments in the modeling of the fate of oil and chemical spills since the 1970s. Much of the modeling relies on high-resolution hydrodynamic models along with transport by currents, diffusion, surface spreading, evaporation, vertical dispersion, and emulsification (Hodgins et al, 1991). Transport by currents is a significant factor in the fate of the trajectory of an oil spill. The surface current provides the majority of the force that spreads a spill and is significantly derived from tidal currents and wind induced surface currents. These two components are directly accessible in near real-time from the TCOON. Current meters have been successfully deployed on DCP's, however, to date water current has not been measured for extensive lengths of time due to high cost of instrumentation and maintenance.

The Oil Spill Prevention and Response Program at the TGLO has purchased SPILLSIM software to model oil spill along the Texas coast. SPILLSIM incorporates an application of a barotropic numerical current model (C2D- denoting currents in two dimensions). The transportation of oil is given by two momentum equations, one for the x-direction (positive to the east) and one for the y-direction (positive to the north). The software requires the following input parameters to drive the oil spill model:

- water current databases (generally not available in Texas)
- Real-time HF radar currents (not available in Texas)
- modeled currents (no reliable models completed for Texas)
- winds
  - -real-time (available from TCOON)
  - -forecasts (available from the National Weather Service)
- geomorphology (available from National Oceanic and Atmospheric Administration)
- commodity properties (available from the oil company)

Without real-time data, the parameters entering the model must be estimated from tidal predictions and predicted wind speed and direction. As Texas coastal bays and estuaries are shallow with very unpredictable tidal regimes, the model is not likely to produce accurate trajectory predictions without accurate real-time data. TCOON has been designed to supply the data needed to drive the spatial components of oil spill trajectory models such as SPILLSIM.

THE GIS USER INTERFACE

In order to capture and display real-time data in an easily readable form, GIS has been used as an interface to the TCOON data. The user interface for RTNET is being developed by the CBI using ARC/INFO Arc Macro Language (AML) to implement a variety of graphical displays.

Once the RTNET interface is initiated, the outline of the Texas coast is displayed. Also displayed are the locations of all RTNET platforms. As part of the initialization of the AML, the current RTNET data file (transferred via the Internet) is also imported into INFO. The user may then zoom into the appropriate area on the Texas coast map. Based on the data in the RTNET file, the AML's also display a vector for each station representing the wind speed and direction at each station. The magnitude of each vector gives the user an indication of the wind speed relative to other gauges in the vicinity and the direction of the vector shows the prevailing wind direction.

Through a series of selection screens, the user is able to choose a particular gauge to 'pop up' a window which will display extended data from the RTNET data file for all sensors and the time in minutes since each reading was taken. The most recent measurements for each sensor are displayed, together with graphs of historical data for water level and air temperature. The water level data also includes the highest water reading for the last 12 hour period. When the user views this extended data screen it is anticipated that they will be able to determine whether the tide is rising or falling. In addition, it is hoped that they may be able to intuitively extrapolate from the historical data graphs the approximate height and time of the next highest tide. This would be of critical importance in determining the size of vessels which could be used in the cleanup operations in shallow bays and estuaries.

The main map screen also includes a field which indicates if more recent data is available. This is highlighted whenever a new RTNET data file is transferred from the CBI. The user then has the option of updating the displayed values by reading in the new data file. All displays are designed to give the user an intuitive feel for both the type of data displayed and also for navigation around the RTNET map.

At the time of this writing, the user interface is in the testing phase and it is expected that some modification will be necessary prior to full implementation of the interface on the TGLO computers. The inclusion of current meters to the TCOON platforms would also provide valuable data to trajectory models. The RTNET data files and interface have been designed to facilitate easy incorporation of additional sensors such as current meters into the network (Garrett and Jeffress, 1992).

According to the US Coast Guard, the highest risk of an oil spill in Texas comes from the collision of vessels carrying crude oil and refined oil products (McHenry, 1991). Much of the crude oil that is refined in Texas is imported from the Middle East, South America, and Africa. Super tankers importing crude have drafts greater than the depths of the ship channels serving the ports in Texas. As a result crude oil is "lightered" from the super tankers onto smaller tankers offshore then ferried to the refineries along the ship channels leading to major ports. The super tanker *Megaborg* was lightering when it blew up off the Texas coast in 1990. Lightering increases the volume of traffic using the ship canals which in turn increases the risk of collision. Access to real-time meteorological and current condition data along the ship channels and the intracoastal waterway may provide useful information for pilots, ships captains, and barge operators to prevent collisions from occurring.

GIS has recently been used in response to oil spill disasters such as the Exxon Valdez (1989), World Prodigy (1989) and the American Trader (1990). This use has generally been confined to the mapping and record keeping of data relating to the spills (August et al 1990, GenaNews 1990). This represents reaction to the spill, rather than the use of GIS as a tool to manage and coordinate the overall spill response. To effectively use GIS for this task requires the establishment of extensive databases of information prior to the actual event. August et al (1990) outlines the types of GIS preparation which should be carried out.

An ideal system would incorporate data on the location and quantity of cleanup equipment and personnel, an inventory of natural resources around the potential spill site and the location of features such as fish hatcheries and desalination plants, etc., which could be threatened by a spill. In addition, the coastline should be indexed based on the environmental 'sensitivity' of the shoreline material to oil inundation. Provision should also made for the inclusion of satellite imagery for monitoring of the spill. Jayko (1990) also identifies the establishment of the data paths through the data synthesis hierarchy, as a major obstacle early in the spill response for the World Prodigy. These data paths should be identified and tested by simulated exercises prior to the spill, rather than during the response.

It is an enormous task to prepare and maintain a large spatial database. The U.S. Coast Guard has considered the establishment of such a database (Jenson, 1990). Many local authorities are also undertaking the compilation of such databases as part of resource mapping for the management of tidal wetlands, bays and estuaries. The preparation of comprehensive oil spill contingency plans which will reside within the same database are another aspect of the effective management of any coastline. It is a mandatory undertaking if the full functionality of GIS to manage an oil spill response and mitigation is to be realized. (Garrett and Jeffress, 1992)

As the oil industry has painfully realized, oil spill concerns do not end with adequate cleanup and mitigation. Companies responsible for oil spills face large compensation litigation actions from those affected by oil spills. GIS offers a useful tool as a means to record the events leading up to, during, and after an oil spill. As long as adequate measures are put in place to ensure spatial and attribute data accuracy, GIS should provide suitable evidence to minimize the litigation of the extent of damage caused by an oil spill. The key to this aspect of GIS use is making sure accuracy of the data used is to a standard acceptable by the courts. This may imply the use of mapping accuracies similar to those used in the production of nautical charts and the verification of attribute data by persons with acceptable expertise. (Jeffress, 1993)

CONCLUSION

There will always be the risk of oil spill as long as the demand for oil based products exist. The antidote to risk is information. GIS offers a means to supply useful information for the prevention, response, mitigation, and litigation of oil spills. The incorporation of real-time data into GIS increases the usefulness of GIS derived information. The Texas Coastal Ocean Observation Network provides real-time environmental data that can be incorporated into oil spill dispersion and fate models as well as GIS, and will add to the information that is needed to reduce the risk and cost of oil spills.

ACKNOWLEDGMENTS

Support for this paper has been provided by the Texas General Land Office through contract No. 92-170R and the Conrad Blucher Institute for Surveying and Science, Corpus Christi State University.

REFERENCES

Anonymous, "Genamap Helps Battle Oil Spill", GenaNews, Vol. 1, No. 1, Genasys II Inc., p 3, 1990.

August P., S. Hale, E. Bishop and E. Sheffer (1990), "GIS and Environmental Disaster Management: World Prodigy Oil Spill", Proc., Oil Spills: Management and Legislative Implications, American Society of Civil Engineers, New York, pp. 366 - 374.

Butler, H. L., R. S. Chapman, B. H. Johnson and L. J. Lower, (1990), "Spill Management Strategy for the Chesapeake Bay", Proc. Oil Spills: Management and Legislative Implications, American Society of Civil Engineers, New York, pp. 325 - 333.

Garrett, M. and G. A. Jeffress (1992), "Incorporation of Real Time Environmental Data Into a GIS for Oil Spill Management and Control", Proc. GIS/LIS Annual Conference, San Jose, California, Vol. 1, pp. 247-255.

Hodgins, D. O., Jian-Hua Huang, D. S. Dunbar, and S. L. M. Hodgins (1991), "A High-Resolution Coupled Hydrodynamic and Oil Spill Modeling System Applied to the Port of Vancouver", Proc. Fourteenth Arctic and Marine Oilspill Program Technical Seminar, Vancouver, Canada.

Jayko, K. (1990), "Predicting the Movement of the World Prodigy Spill", Proc., Oil Spills: Management and Legislative Implications, American Society of Civil Engineers, New York, pp. 351 - 365.

Jeffress, G. A. (1993), "Real-Time Environmental Data for Oil Spill Response", Proc. of The 1993 ESRI Users Conference, In press.

Jeffress, G. A. (1991), "Next Generation Water Level Measurement for the Texas Coast", Proc. of The Second Australasian Hydrographic Symposium, The Hydrographic Society, pp. 407 - 415.

Jenson D. S. (1990), "Coast Guard Oil Spill Response Research and Development", Marine Technology Society Journal, Vol. 24 No., pp. 39 - 42.

Liebermann, T. D. and J. C. Ciegler (1992), "Interactive Display and Query of Near Real Time Data from a Hydrologic Alert Network", Proc., 12th Annual ESRI User Conference, ESRI, pp. 139 - 149.

McHenry, Lt. Comdr. W. (1991), Commander of the Office of Marine Safety, US Coast Guard, Corpus Christi, Texas, personal communication.

Managing Dryland Salinisation with an Integrated G.I.S/E.S. System.

*S.D.Kirkby and F. Kurzel**
Department of Geography
Department of Computer Science*
University of Adelaide
Adelaide 5005
Australia
email skirkby@fenner.geog.adelaide.edu.au

Abstract

This paper considers the development of a system which integrates a Geographic Information System (G.I.S.) an Expert System (E.S.) and a relational database management system (RDBMS) to identify and manage secondary salinisation. There are two issues addressed. The identification of known saline areas, potential saline areas (discharge zones) and groundwater recharge zones utilising heuristic knowledge and the development of a user friendly interface which allows a user to both interrogate the decision making process and update the data within the RDBMS. The spatial occurrence of dryland salinisation ensures that appropriate rehabilitative management techniques require implementation by numerous landowners within a region. To ensure management strategies are implemented it is imperative that a landowner is fully informed of the decision making process to classify salinity potential.

Introduction

Dryland salinisation, currently estimated at 225,000 hectares is the worst land degradation problem existing in South Australia (SDSC,1990). To ameliorate secondary salinisation, appropriate high water usage management strategies require implementation at both groundwater recharge and discharge areas. As yet, no physical based environmental model exists which can identify these areas at a regional scale as the complexity of input parameters varies widely. Consequently, potential areas of groundwater discharge and recharge are difficult to identify resulting in few proactive rehabilitation strategies being implemented.

Conversely, heuristic information for secondary salinisation is well documented (Jolly, 1988). By systematically formulating this knowledge into a series of if/then production rules within an Expert System (E.S.), the conditions conducive to the development of groundwater discharge and existence of recharge may be determined. To spatially identify the specific parameters which lead to the development of dryland salinisation a Geographic Information System (G.I.S.) database was formulated. By combining the E.S. and the G.I.S. through a common Relational Database Management System (RDBMS), a method has been developed to map the spatial occurrence of existing and potential groundwater discharge and recharge areas.

The Nature of Salinity

The cause of secondary salinisation is the addition of groundwater via groundwater recharge in excess of that required to maintain a balance with water use and losses. It results in rising groundwater which mobilise the salts towards the soil surface. The water recharge (net downward movement of water) / discharge (net upward movement of water) system is the factor that controls salinisation and it is determined by the spatially variable geological, geomorphological, pedalogical, hydrological and landform characteristics of an area. It is important to note that groundwater discharge is dependent on the amount of groundwater recharge i.e. an increase in groundwater recharge results in an increase in groundwater discharge.

Some groundwater recharge is likely to occur over the whole area of a catchment, however the volume of recharge per unit area will vary greatly in both time and space. Dyson (1983) identified 30% of the Axe Creek catchment in Bendigo Victoria which supplied 83% of the groundwater recharge, whilst Cook et al.,(1989) identified 20% of the Borrika catchment in South Australia which supplied 41% of the total recharge. These preferred zones of recharge are due to the soil, geologic, land cover and topographic characteristics of the catchment.

Identifying the location of groundwater recharge areas is difficult (Williamson, 1990). Cook (pers. comm., 1992) states that when using empirical models, such as Modflow, 20 to 40 holes on average need to be drilled in any one catchment to determine the major recharge zones. As drilling is labour intensive, slow and expensive both Cook (pers. comm.,1992) and Morris and Thomson (1983), acknowledge that it will be some years before extensive recharge data is available for the major salt affected catchments of Southern Australia. Dyson (1990) identified three major types of groundwater systems. Firstly local systems occur when the groundwater cannot flow over surface water catchment boundaries due to the geologic and soil conditions. The salinity occurring in these areas may directly be related to the land management within the catchment. Secondly, regional groundwater systems typically occur in sedimentary basins where the groundwater moves freely through the deeper perennial aquifer recharging the groundwater table leading to its eventual rise and the subsequent saline seepage. Management of these systems is difficult as it requires a coordinated effort over a region where recharge is occurring. The third system is the intermediate groundwater system where water flows across catchment boundaries and throughout large areas, but is not regional as the process occurs over tens of kilometres rather than hundreds of kilometres. The management of this groundwater system is also very difficult due to the need to implement rehabilitation programmes over the whole area.

The discharge of the saline water from the groundwater system is by either flow into a stream or by capillary rise and subsequent evaporation from the soil surface. With active salinisation in a saline seepage area the salt levels are expected to be higher near the surface where the soluble salt load is deposited. Where upward movement is not occurring (or occurring at a very slow rate) infiltration of rainwater tends to leach soluble salts from the surface into the subsoil increasing the salt content with depth. There is an established relationship between discharge areas and the vegetation cover. Increasing soil salinity in the discharge zone subjects the vegetation cover to increasing salt stress, suppressing plant growth conditions, leading to the establishment of salt tolerant halophytic species and subsequently senescence of the halophytes. Removal of the vegetative cover enhances the possibility of significant soil erosion due to the dispersive nature of sodium enriched soils (Peck, 1978).

Groundwater discharge areas are easily discerned by remote sensing utilising vegetation as a surrogate variable (Chaturvedi et al.,1983; Hick and Russell, 1990; Kirkby, 1993). Salinisation trends have been quantified but not the potential for salinisation to occur. Attempts to identify potential groundwater discharge have been made by Jackson and O'Neil (1987), Reutov (1989) and Singh et al.(1990) using radar imagery and Huntley (1978), Price (1980) and Jupp et al. (1990) using thermal imagery.

Remote sensing techniques have been utilised to identify groundwater recharge areas, but as there is no one surrogate which indicates groundwater recharge, results have not been favourable (Mackenzie et al.,1990). One approach which attained some success was by Hill (1992), who integrated land system data with a stratified Landsat TM image. By assuming the influence of soil, geologic and topographic characteristics on groundwater recharge Hill (1992) was able to identify potential areas of groundwater recharge. This basic premise has been adopted by this paper.

Current Management Strategies

Management strategies are short term and/or long term. Short term strategies deal with the saline seepage areas where the groundwaters are discharging at the surface. These strategies deal with the problem once it has developed, they do not attempt to prevent it from occurring. The reasons for this relate to the relative ease by which groundwater discharge may be identified.

Traditionally, the removal of trees has been recognised as the main cause of secondary salinity and hence most short term strategies have concentrated on planting appropriate tree species. Trees draw down the watertable, stop capillary rise and allow the leaching of surface salts. In this role they are biological pumps. Studies have shown that water tables can be significantly lowered by the use of tree species. Webster (1983) noted that planting of trees in the Wimmera-Mallee region successfully lowered the water table in some cases by up to 2 metres. Unfortunately tree planting in discharge zones reduces the volume of surface runoff as well as groundwater discharge. Thus water which would have normally run-off, accumulates in the discharge zone further compounding the problem (Williamson, 1990).

Eventually it was realised that as most groundwater discharge zones occur on privately owned land, rehabilitative strategies would only be implemented if they were economically viable (Dyson,1990). This resulted in research concentrating on identifying high-water use salt tolerant fodder crops to be planted on discharge zones.

Rehabilitation of potential groundwater discharge areas would benefit from the same short term strategies, but as no method was available to identify these critical areas no management strategies were implemented.

Long term management strategies aim to stop excess recharge within the landscape entering into the hydrological system. They are deemed long term strategies as the lag factor of groundwater recharge ensures that even if groundwater recharge was halted there would still be water moving through the hydrological system. A problem with the implementation of these management strategies relates to the inability to identify groundwater recharge zones, as already considered, recharge areas vary from metres to tens of kilometres from the relevant discharge areas. Landowners in a region have had no real opportunity to implement preventative rehabilitation strategies. Without this information they are unaware of the implications of their management on land owners in other areas.

Proposed Management System

Description of Study Area

The study area is a 10 x 12 square kilometre region near Jamestown in the mid-north of South Australia. It is approximately 200 kilometres NNW of Adelaide. Contained within the study area is a catchment which has been subjected to intensive field study since 1990 by the State Dryland Salinity Committee. The areas of groundwater recharge and discharge identified by this project will be compared to the results derived by the State Dryland Salinity Committee.

Hardware Software Components

The software being utilised in this system is the Knowledgeworks Expert System shell, the Oracle RDBMS and the Arc/Info G.I.S. The model is being implemented on a Sun IPX workstation.

An Overview

To make all land owners in a region more accountable for their management practices an interactive classification method for dryland salinity has been developed. As an agricultural extension tool, it is a software package designed to be used in the regional agricultural field offices by the region's land owners and agricultural extension officers.

Traditional parametric/G.I.S land classification approaches utilise both G.I.S digital maps, hardcopy thematic maps and handbooks to provide an explanation for the decision making process to the user. These methods tend not to convey understanding of the decision making process to the user. A farmer who wants to know the rationale for why an area was classified a certain type will more likely than not, have no concept of how to operate a G.I.S., thus the information contained within the database and the method utilised to analyse the data will not be transparent to them. Similarly the methodology utilised to formulate a hardcopy land class map is documented in an accompanying handbook, but it is usually documented in a manner which conveys understanding only to an individual familiar with the knowledge domain. It usually does not consider the final user who will implement this information. It is assumed that if a user has an understanding of why a section of land was classified as a certain type, they will be more confident about the information and therefore more likely to utilise it in their decision making process.

An Expert System (E.S.) has been developed to format qualitative knowledge known about salinisation into a series of if/then production rules which are used to interrogate data within a Geographic Information System (G.I.S). A resultant salinity classification map, indicating areas of existing and potential discharge and recharge, is produced and presented to the user as a graphic output on the computer screen. The explanation facilities of the E.S. are utilised to ascertain both the rationale for the classification of an area and the associated management strategy. The E.S. also allows the user to interrogate the G.I.S. database and if required alter the information.

Expert Rules

The formulation of the production rules was an exhaustive process involving a series of taped interviews with five South Australian experts. Initially a flat rule base structure containing 139 rules was developed e.g.

If Soil = 13 and
If Geology = 7 and
If Groundwater <= 3 and
If Elevation <= 2 and
If Slope <= 4
--->
Then land classification is "low potential discharge".

These rules were condensed into 15 membership rules, an analogy is the hierarchical integrated land classification approach. Six general salinity classes were identified;

- high potential groundwater recharge
- medium potential groundwater recharge
- low potential groundwater recharge
- low potential groundwater discharge
- high potential groundwater discharge
- existing groundwater discharge.

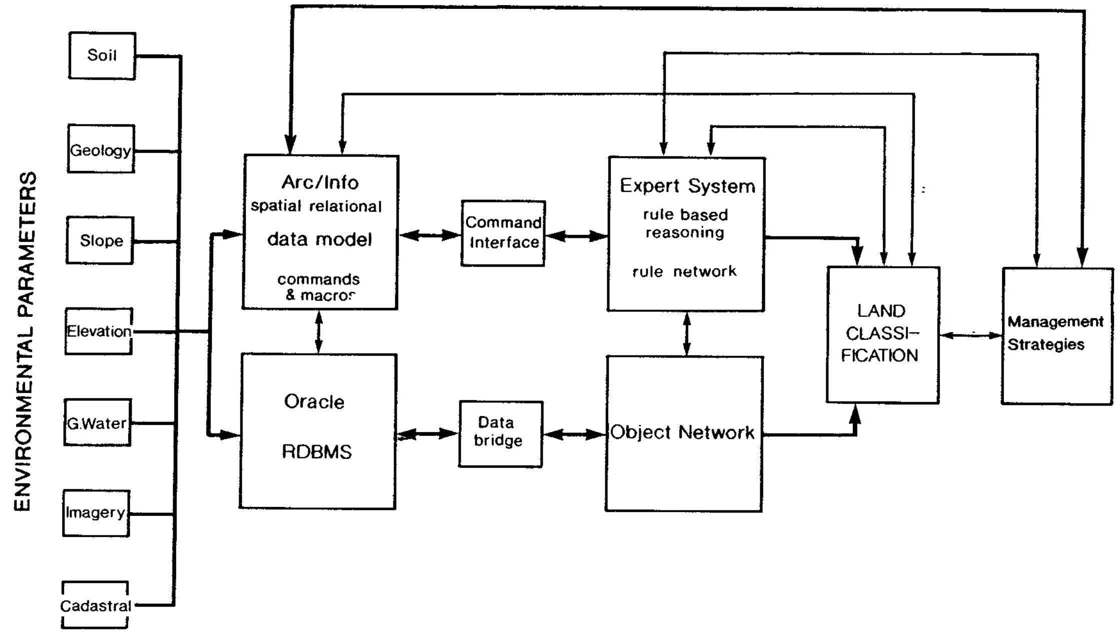
Data Layers

Data was collected for the study area; soil, geology, slope, elevation, groundwater, drainage lines, cadastral, roads, spot heights, contours and a Landsat TM 7 band scene. The soil and geology data layers were digitized from maps provided by the South Australian Department of Agriculture and then converted to a raster data structure by the Grid module in the G.I.S. The slope and elevation data was generated by the TIN module and converted to a raster data structure within the G.I.S. It was derived from digital contour and spot height data provided by the South Australian Department of Lands. The groundwater data was collected from wells surveyed by the South Australian Department of Agriculture. This data was interpolated using the krigging module within the G.I.S and also converted to raster data structures. The Landsat TM scene was manipulated by two spectral indices within the MicroBRIAN image processing system and then transferred directly to the raster G.I.S. module. All raster coverages were transformed to the same AMG coordinates and resampled to 30x30m gridcell size. All data layers were inputted as tables into the RDBMS and subsequently related to the G.I.S tables.

Integration

Upon completion the rules within the E.S. were applied to soil, geologic, topographic (slope/elevation), groundwater and Landsat TM data held within the Oracle RDBMS. The resultant salinity classifications result were written to a new column within the RDBMS. The data within this column could immediately be displayed by the raster G.I.S. The model is considered in Figure 1.

Figure 1: **System Diagram**



It is important to note that the environmental data held within the RDBMS retains the original topology from the raster G.I.S. Thus the specific column, row location of each gridcell is retained within the RDBMS. This enables flexible manipulation of the data by all three software packages.

Interrogating the Salinity Class Map

The user is offered the option of interrogating the land classification map produced by the E.S. To commence this process the user simply selects the area classified from the E.S. This automatically initiates the G.I.S. and draws the salinity classification map (Figure2). This map can be overlayed with any of the vector data layers (e.g. cadastre boundaries) to aid with interpretation. The user may interrogate the map at three scales, gridcell (30x30m), cadastre unit or by regional class. The appropriate level of interrogation is chosen by the user from a menu displayed on the G.I.S. Once selected the relevant information is written to an ascii file which can be read by the E.S. and information regarding the decision making process displayed.

At the gridcell scale an explanation of the decision making process using the flat rule base (139 rules) is provided on a popup screen with additional management suggestions i.e. on soil type A2 with geology type B1 on slope C2 with elevation D1 the appropriate high water use management strategy for salinity class 5 would be E2. A pix-map of a scanned photograph typical of the relevant salinity class also prints on the screen. At the regional and cadastre scales the membership rules are utilised to explain the salinity class results for selected regions. Additional information provided includes an assessment of the total area classified by salinity class.

Updating Databases

Many landholders have valuable knowledge of local environmental parameters, such as depth to groundwater and it is important to capture this information. Updating the database occurs at the gridcell scale via a pop-up screen from the E.S. or by editing an Arc/Info point coverage.

The end user may select from the E.S, a map depicting any parameter, i.e. soil. Once the map is displayed they select the relevant area and the E.S. subsequently displays the attribute information, i.e. clay content, for the selected soil type. Options based on different attribute information are provided with the choice for a different soil type. Most data layers can theoretically be updated via the E.S. Groundwater, slope and elevation may only be altered from the G.I.S. These raster data layers are interpolated from point data to a grid value. By providing a menu for the editing module within the G.I.S. the point data may be updated by the user. Interpolation from the updated point information follows automatically.

Upon completion of the updating process the user is asked by the E.S. if they wish to commit all changes to the database. If "no", all updated information is cleared, if "yes", the salinity land classification procedure is initiated and a new salinity class map based on the new information is produced. At this stage of the system development the run time for the E.S. processing the data within the RDBMS is approximately 4 hours. As the system is designed to have interactive capabilities with the user this is not desirable.

A feature of the updating process is the use of the archiving facility within the RDBMS. All changed values for each gridcell are stored in archive files which can be accessed at any later date. This facility allows temporal analysis of environmental change.

Results of Work to Date

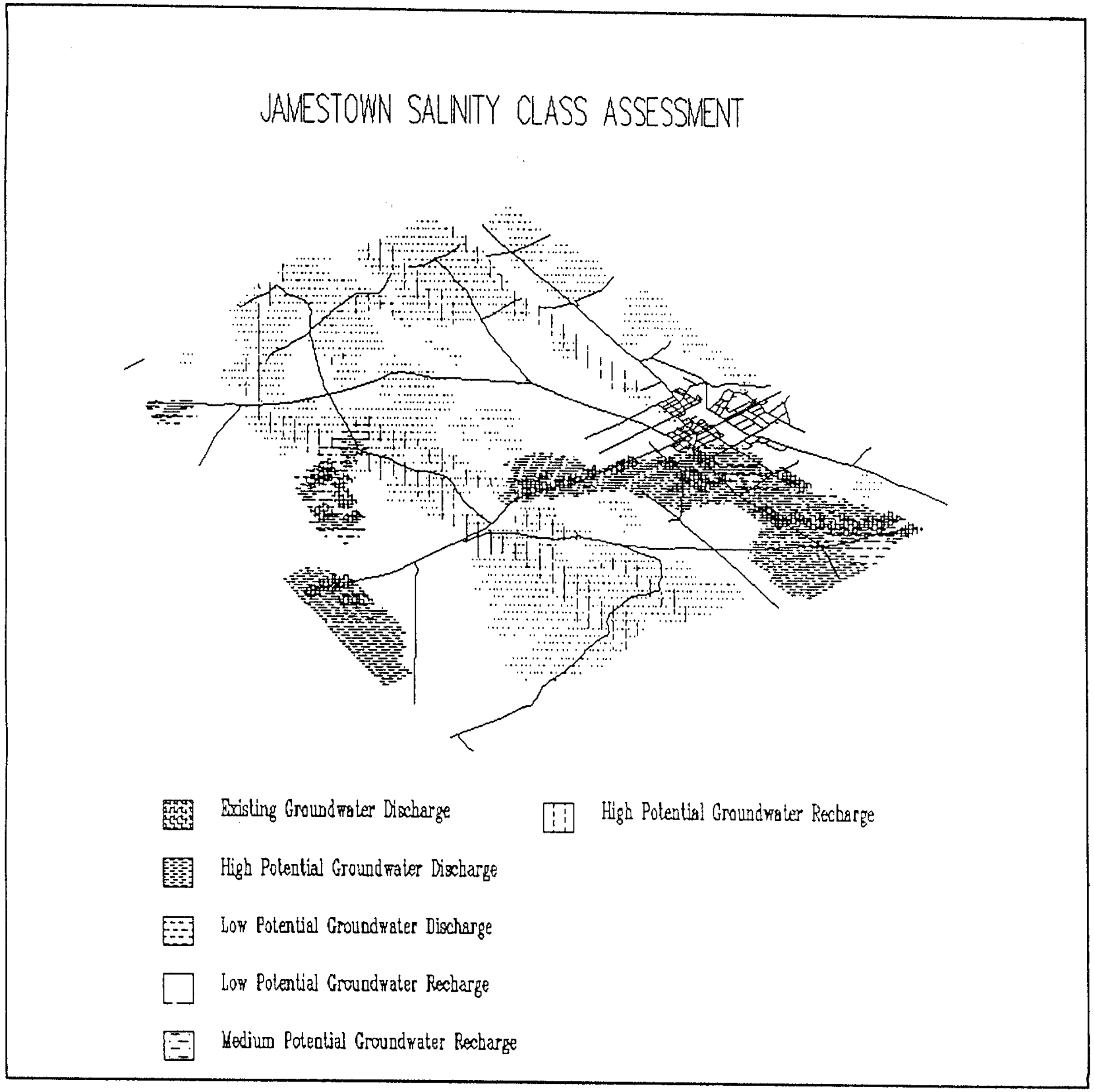
The final salinity classification of the Jamestown region, although comparing favourably to the results produced by the State Dryland Salinity Committee, has not been subjected to any statistical evaluation (Figure 2). At this stage the interrogative component of the system has been finalised. The updating facilities are nearing completion but they require further work.

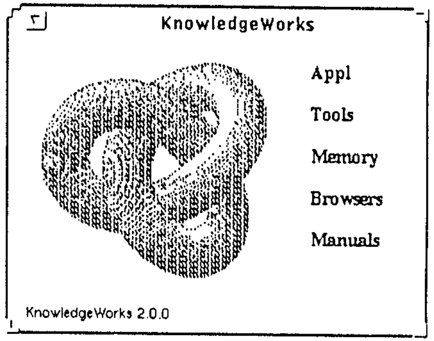
Problems

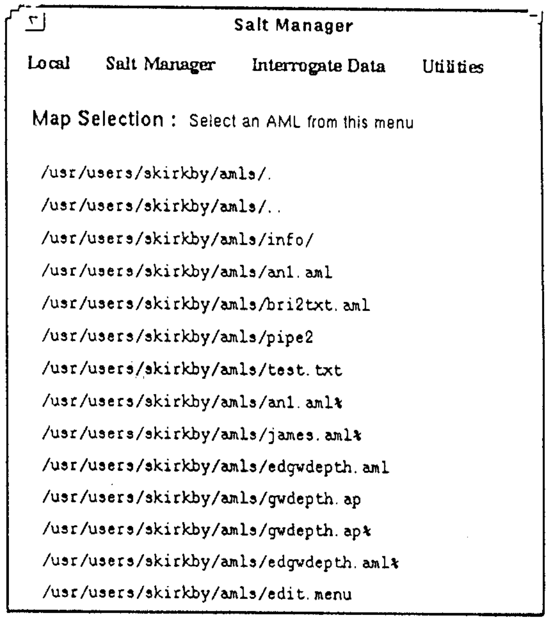
The raster model requires examination to determine errors. The rectification and registration processes of the 12 data layers and the vector/raster conversion of the soil and geology data layers are two such areas where this type of error may have been incorporated. Other sources of potential error are the interpolation processes utilised to develop the elevation, slope and groundwater coverages. A conceptual problem which requires consideration is the data structure of the raster data. The gridding of the landscape reduces the elementary environmental components into a continuous layer of cells. As all operations within the system are conceptualised on the grid rather than on the environmental fields, the model becomes divorced from reality. This concept needs to be addressed.

Conclusion

The salinity management system provides three potential benefits to landholders in salt affected regions of South Australia. It is a means by which existing and potential groundwater recharge and discharge areas may be identified, thereby indicating the extent of the salinity problem. It offers management advice from an interactive computer screen which provides textual and graphic information. It allows information held by landowners to be incorporated into the salinity class decision making process. Once the system has been fully developed it will be trialled by the CSIRO and the South Australian Department of Primary Industries in other salt affected regions.


JAMESTOWN SALINITY CLASS ASSESSMENT
Existing Groundwater Discharge
High Potential Groundwater Discharge
Low Potential Groundwater Discharge
Low Potential Groundwater Recharge
Medium Potential Groundwater Recharge
High Potential Groundwater Recharge


KnowledgeWorks
Appl
Tools
Memory
Browsers
Manuals
KnowledgeWorks 2.0.0


Salt Manager
Local
Salt Manager
Interrogate Data
Utilities
Map Selection : Select an AML from this menu
/usr/users/skirkby/amls/.
/usr/users/skirkby/amls/..
/usr/users/skirkby/amls/info/
/usr/users/skirkby/amls/anl.aml
/usr/users/skirkby/amls/bri2txt.aml
/usr/users/skirkby/amls/pipe2
/usr/users/skirkby/amls/test.txt
/usr/users/skirkby/amls/anl.aml%
/usr/users/skirkby/amls/james.aml%
/usr/users/skirkby/amls/edgwdepth.aml
/usr/users/skirkby/amls/gwdepth.ap
/usr/users/skirkby/amls/gwdepth.ap%
/usr/users/skirkby/amls/edgwdepth.aml%
/usr/users/skirkby/amls/edit.menu

Acknowledgements

We would like to thank Errol Bamford, Dr. Peter Eklund and Dr. Maureen Longmore from the University of Adelaide for their incisive comments during the development stage of this project. The assistance of Cameron Humphries from the University of Adelaide and Jason Trenouth from Harlequin Ltd has also been invaluable!! Special thanks must go to Petra Heeg and Anne Kurzel.

References

Chaturvedi, L., K.R. Carver, J.C. Harlan, G.D. Hancock, F.V. Small, K.J. Dalstead (1983), "Multispectral Remote Sensing of Saline Seeps", in IEEE Transactions on Geoscience and Remote Sensing, Vol. GE-21, No.3, pp 239-251.

Cook, P.G., G.R. Walker, I.D. Jolly (1989), "Spatial Variability of Groundwater Recharge in a Semi-Arid Region", in Journal of Hydrology, Vol. 111, pp 195-212.

Dyson P.R. (1983), "Dryland Salting and Groundwater Discharge in the Victorian Uplands", in Proceedings Royal Society of Victoria, Vol. 95, No.3, pp 113-116.

Dyson P.R. (1990), "Dryland Salinity in South Eastern Australia. An Overview of Salinity Processes", in Management of Soil Salinity in South East Australia, pp 231-240, ed. by Australian Society for Soil Science Inc.

Hick P.T. and W.G.R. Russell, "Some spectral considerations for Remote Sensing of Soil Salinity" in Australian Journal of Soil Research, Vol. 28 No. 3.

Hill S., (1992), "Recharge Mapping in the Avoca Catchment Using Remote Sensing and Geographic Information Systems" in 6th Australasian Remote Sensing Conference, Wellington N.Z. pp 360-368.

Huntley D., (1978), "On the Detection of Shallow Aquifers Using Thermal Infrared Imagery", Water Resources Research, Vol 14, No. 6, pp 1075-1083.

Jackson T.J., and P.E. O'Neil (1987), "Salinity Effects on the Microwave Emission of Soils", in IEEE Transactions on Geoscience and Remote Sensing, Vol GE-25, No.2, pp 214-220.

Jolly I.D., (1988), Non-irrigated salinity in South Australia: Options for Treatment, South Australian Department of Agriculture, Technical Report No. 124.

Jupp, D.L.B., J. Walker, J.D. Kalma, T. McVicar (1990), "Using Thermal Remote Sensing to Monitor Land Degradation and Salinisation in the Murray-Darling Basin of Australia", in 23rd International Symposium on Remote Sensing of the Environment, Bangkok, Thailand, April 18-25.

Kirkby, S.D. (1993), "Monitoring Saline Discharge Areas: Farrell Flat a case study" in Australian Geographical Studies, Vol.31, No.1, pp.62-69.

Mackenzie M., G. Bellamy, D. Fraser, G. Ellis (1990), "Mapping Dryland Salinity Recharge Areas using EM conductivity measurements and Airbourne Scanner Imagery", in 5th Australian Soil Conservation Conference, Remote Sensing and GIS Workshop, Perth W.A.

Morris, J.D. and L.A.J. Thomson (1983), "The Role of Trees in Dryland Salinity Control", in Proceedings Royal Society of Victoria, Vol. 95, No. 3, pp 123-131.

Peck A.J., (1978), "Salinisation of non-irrigated Soils and Associated Streams: A review" in Australian Journal of Soil Resources, Vol.16, pp 157-168.

Price J.C., (1980), "The potential of Remotely Sensed Thermal Infrared Data to Infer Surface Soil Moisture and Evaporation" in Water Resources Research, Vol.16, No.4, pp 787-795.

Reutov E.A. (1989), "Microwave Radiometry of Saline Soils" in IGARSS '89 12th Canadian Symposium on Remote Sensing, Vol 5.

Singh R.P., V.Kumar, S.K. Srivastav (1990), "Use of microwave remote sensing in salinity estimation" in International Journal of Remote Sensing, Vol.11, No.2, pp.321-330.

State Dryland Salinity Committee (SDSC), (1990), Technical Strategy to Address Dryland Salinity in South Australia, S.A. Department of Agriculture.

Webster A, (1983), "Intercepting Channel Seepage in the Wimmera-Mallee: the use of trees and tall wheat grass" in Water Talk

Williamson, D.R., (1990), Salinity - An Old Environmental Problem, CSIRO Technical Memorandum 90/7, Division of Water Resources.

CAD software, GIS software and Database Management Systems (DBMS). However, the term GIS can be somewhat understood by looking at the technology, data and applications that have been classified under the GIS banner. These will be further described below along with the implications for using GPS derived data in GIS databases.

GPS for GIS

GPS Accuracy

The GPS specifications of importance for GIS are the positioning capabilities of the system. Essentially, there is a trade-off between price and accuracy in both GPS receiver design and also field operations and after survey data processing. The various scenarios are described in Figure 1. With a lot of money spent on receivers, a lot of time in the field and also a lot of time processing the data, extremely precise and accurate measurements of baselines can be made; as accurate as any long distance measurements that have ever been made. Here we refer to a baseline as being the three dimensional vector between two points. On the other hand, for about $1000 a handheld GPS receiver can give positions in real time accurate to about ±100 m.

It is important for GIS applications that we understand the different modes of GPS operations and the results that can be achieved. We should remember that 100m accuracy does mean that my displayed position, which is updated every second or so, can change very rapidly even whilst I am standing still. It is interesting to read in some GIS literature that this concept is often quite confusing!

Absolute Position		**Accuracy**
Navigation (real time)		
	SPS with SA	100 m
	SPS without SA	20 m
	PPS with A-S	8 m
Relative position		
Differential Code (post processed; real time)		<10m
Carrier-smoothed code (post processed; real time)		1m
Kinematic carrier phase (post processed; real time)		<10cm
rapid static carrier phase	(minutes)	cm
static carrier phase	(Hours)	ppm

SPS...standard positioning service — PPS...precise positioning service
SA...Selective Availability — A-S anti-spoofing

Figure 1
GPS Positioning Modes

GPS for GIS

E.G.MASTERS
C.RIZOS
B.HIRSCH

School of Surveying
University of New South Wales
PO Box 1
Kensington 2033
NSW

ABSTRACT

Global Positioning System (GPS) technology is rapidly being adopted as a "real world" digitising device linked to Geographical Information Systems (GIS). This paper describes aspects of the GPS and GIS which must be considered when these technologies are being interfaced. These aspects include transformations between coordinate systems, GPS and GIS accuracy issues and GIS data structures. Finally a few comments are made on the future directions that the link between GPS and GIS will take.

INTRODUCTION

What is GPS?

The Global Positioning System or NAVSTAR navigation system, often referred to as simply GPS, is a United States Department of Defense, satellite-based navigation system. The system is designed to be 24 hour, operating in all weather and having global coverage. These requirements, obviously of great utility to the military, have also essentially been made available to the global civilian community at little cost. The general specifications for the GPS system, the satellites, the control segment, etc., have been described in many well-known publications and will therefore not be elaborated upon here. For a simple and easy to read explanation see McElroy (1992).

The advent of relatively inexpensive and easy-to-use GPS receivers has led to an explosion in the applications and number of users of position information. GPS provides an excellent tool for locating all types of features and themes and mapping them. Unfortunately, the ease of use of such technology belies some of the complexities of coordinate systems and errors that can occur when using such technology, especially when data is being integrated with other databases. A few of these issues will be discussed in this paper.

What is GIS?

GIS is a term that means many things to many people. It is difficult, if not impossible, to tie down and define exactly. As the whole area of spatial data handling is undergoing rapid development and evolution, GIS is becoming a generic term to cover technology and databases which deal with any locational or spatial data. As we live in and manage a spatial world many technologies and databases can therefore be categorised under the term GIS, even ones which may not normally be considered as part of the geography discipline or have the conventional GIS functions. An example of the blurring of technologies can now be seen in the linking of

GPS Coordinate Systems

One of the strengths of GIS technology is to integrate disparate data sets to aid decision making. GPS technology can aid the integration process by providing a uniform coordinate system for base mapping, but the underlying principles of coordinate systems used both for GPS and mapping in general must be understood correctly so that proper integration can be undertaken. Most of the GIS and GPS literature, being of non-Australian origin, does not describe these situations for Australian mapping very well.

GPS receivers will usually be capable of giving 3 dimensional positions, velocity and time. However, great care should be exercised in using the positions. Standard GPS coordinates are with respect to the World Geodetic System 1984 (WGS84). This coordinate system is based on one specific definition of the centre of the Earth and the location of zero longitude. Most conventional geodetic datums have no definitive relationship to WGS84. Hence, datums like AGD66 and AGD84, which are used for most mapping in Australia, are over 100m different to the coordinates that GPS receivers calculate. This systematic error is in addition to the accuracies given in Figure 1. Most GPS receivers will provide transformations between these geodetic datums, but they will always only be approximations and hence be fine for some applications and not others.

The heights given by the GPS receivers must be treated separately from the horizontal component of position. GPS heights are determined with respect to a geometric figure called the ellipsoid and can be very different to heights above sea level as shown on most maps. We should also note that GPS determined heights are generally less accurate than horizontal components of position. The accurate conversion from ellipsoidal heights to mean sea level heights is an extremely complex process. Some receivers will have approximate corrections for these differences, the accuracy of which must be matched to the accuracy requirements of any specific application. Also, our sense of height is more acute than horizontal position. It can be rather confusing to the uninitiated when GPS positions occasionally seem to be obviously wrong, even though they are within the accuracy specifications of GPS.

An important point to remember with all these corrections is that sometimes GPS receivers will display coordinates corrected to local geodetic datums, but only WGS84 coordinates may be stored in files and transferred to other software, like GIS packages. GPS receivers vary in their capabilities for recording and processing data. Most manufacturers provide receivers with a range of capabilities, though there are fewer manufacturers of the more specialised, higher priced geodetic surveying receivers than the basic navigation receivers. Price will generally affect the number of functions and display options on any receiver. Some receivers will only display or output data directly to a communications port. Others will store data in internal memory and permit certain amounts of processing to be done. Many receivers are very similar, but the choice of receiver really needs to be matched to each individual application.

Needless to say, apart from the minor difficulties associated with using the technology, GPS has many benefits for GIS users and is rapidly being adopted for many applications. Why is GPS technology being so readily adopted as against other navigation and positioning technologies? For many applications

- it is cost effective,
- it is easy to use, not requiring experts to determine precise locations,
- it is available in remote locations,
- it suits the accuracy requirements of many applications,
- it provides 3 dimensional positions and it also provides time and velocity.

GIS APPLICATIONS and DATABASES

Definition by GIS Applications

GIS applications have historically revolved around decision making processes which required information to be derived from maps. For example, the decision making process of planners requires many themes of data to be overlayed; elevations, slopes, aspects, vegetation cover, soil types, land use, etc. Intersected overlays of these themes are then used to derive the optimum location for planned land uses. Of course, the potential number of applications and associated functionality is as large as there are decision makers and also the types of data collected.

Some typical applications for GIS technology include: land use planning, natural resource management, environmental monitoring, transport planning, facilities management and property management. All of these themes have in common the management of some type of resource which has a spatial component, either at the planning stage or at the operational management stage. The common types of data stored for these applications on GIS systems are cadastral mapping, topographic mapping, water, sewer, drainage, electricity and telephone facilities.

Definition by GIS Technology

An alternative definition of GIS is to limit it to refer to only the technology. A fundamental requirement of GIS technology is to be able to handle map-type data, integrate and analyse such data and produce reports for decision making. GIS technology usually comprises an input subsystem, a data storage and data management subsystem, a manipulation and analysis subsystem and a reporting subsystem (see Marble,1990).

The difficulty here is to separate the functions of GIS from those of other similar types of packages like CAD and other engineering software, which is easier said than done. Of course, it is possible to make a living out of matching software packages like these with the business requirements of various organisations. Also the packages are becoming very similar as vendors quickly adopt the successful functions of competitor's software.

GIS software packages vary as much in function as the applications they support. However, it is possible to define a basic set of functions that could come under the banner of GIS; for example, map management, polygon overlay, network analysis and buffer analysis to name a few.

Spatial Data Types

It is somewhat simpler to define the fundamental data types that are stored in GIS systems. These are described in great detail in various publications (eg. Burrough,1986; Peuquet,1990; STDS, 1988). However, the data types basically can be considered as points, lines, polygons, grids and images all linked to attribute data (see Figure 2). GIS systems are traditionally 2 dimensional in nature, though many systems do cater for digital terrain models, which will be integrated with the other data to varying degrees.

One of the strengths of GIS is its ability to manage and integrate diverse types of information and yet this is one area where the technology is most likely to be misused. Successful integration of disparate data sets requires consistency between coordinate systems and spatial errors. GPS provides a technology that will bring consistency to both these factors.

GPS as a positioning device will interact with GIS at the data input subsystem and also when any managed resource needs to be located in the field. In either case an understanding of the data input subsystem and data storage subsystem is required to link GPS and GIS together.

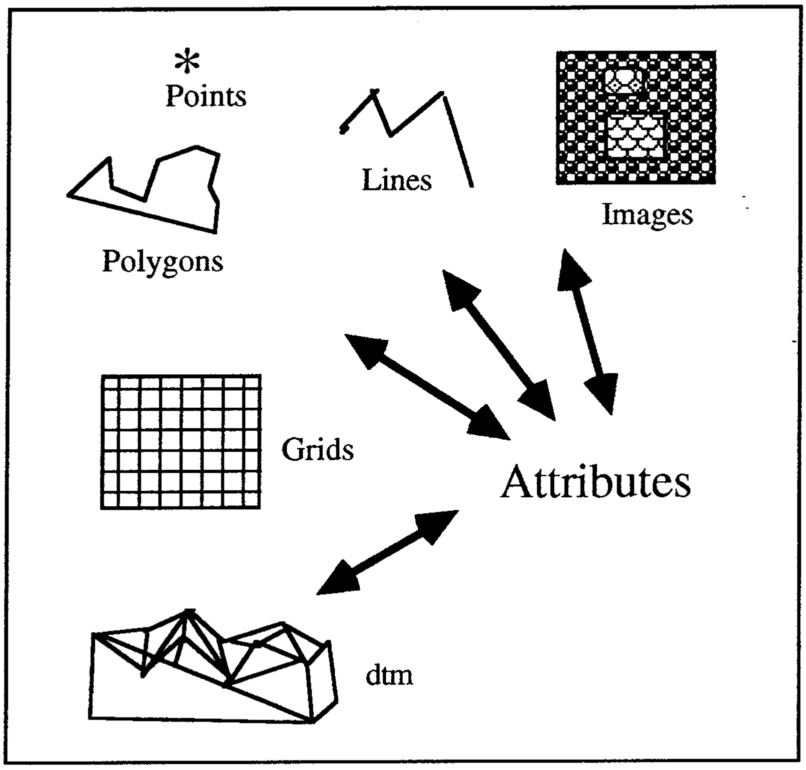


Figure 2
Fundamental GIS Data

The data for most current GIS systems has been digitised from existing map series using tablet digitisers. The inherent problems with this process is that the accuracy of the data is restricted to the scale of the mapping and the data is only as up-to-date as the base mapping. Efficiencies in digitising maps can be gained in some cases by scanning. However, the fundamental problems with using maps as the original data source still remain.

GIS technology is now widely used in the community. However, the typical growth rates being seen for the industry is usually estimated to be around 30% per year, even in recessionary times. This growth rate is being driven by the concerns of our society to manage all types of resources more efficiently. GIS is usually seen to be the enabling technology that will enable this optimisation of management procedures to occur.

WHERE IS GPS USED FOR GIS?

The coordinates adopted for any base mapping which traditionally has been digitised for GIS, will be developed on some geodetic reference system, which in most cases will now be measured using GPS equipment. Most surveys for photogrammetric and map control are now most economically performed with GPS. As a basic geodetic and surveying tool, GPS will therefore always be part of the positioning process for GIS systems.

In remote sensing applications GPS will be used more regularly to locate features as control points to register and scale images to map grids. GPS will also be used to locate features on images in the field for ground truthing or checking the quality of image classifications. However, there is no direct interaction between GIS and GPS at this level, though this could change as the initial map making process becomes more directly linked to GIS databases. For example, it is fairly common now for surveying software and photogrammetric software to provide data directly to GIS software.

There are many documented articles of GPS being used in GIS (see most issues of GPS World and GIS World). These include providing base survey control to coordinate cadastral boundaries, mapping productivity on farms, managing forest fires, managing railway stock, mapping street networks, road asset management, and intelligent vehicle highway systems (IVHS).

Most GIS applications have used GPS as a type of off-line digitiser to provide the locational information for some resource. In these cases, the GPS data has been collected in the field, stored in a receiver or data collector along with attributes describing the features, returned to the office and then transferred to the GIS software and into a database. This type of operation puts few technological requirements on the GPS receiver or GIS database. GPS is basically a non-interactive digitising system in this mode.

Some of the transport and fleet management applications require real time links between the GPS receiver and the spatial database. Many of these navigational type applications require position to be plotted in real time on an electronic map. This process is in itself not difficult to develop, though obtaining the required accuracies from GPS may sometimes be a little difficult. The electronic map in turn enables optimised route navigation to be performed (eg. Rupert,1991). The real time GPS links to other databases in these cases also provides access to other spatial data such as traffic information, which can be also used on line to further optimise route navigation. In these cases the link can be two way as well. Traffic conditions and route information collected at each vehicle can be transferred immediately back to the central traffic management database. In this type of scenario, GPS is totally integrated with the GIS system, somewhat like a mouse and cursor rather than an off-line digitiser.

WHAT ARE THE ISSUES FOR INTEGRATING GPS and GIS?

Providing real-time links between GPS receivers and GIS databases obviously provides the major technological challenge for future GPS/GIS developments. However, for any GPS/GIS application a few issues arise for the use of GPS.

* **Can GPS position data be structured to suit the data structures of GIS?**

  GIS is fundamentally about modelling real world features. GIS data objects are usually built up from points as the fundamental unit. A point is defined by its coordinate location. Lines are defined as strings of vertices or points. Polygons are strings of lines. Features are combinations of points, lines and polygons. All features and objects have associated attributes, which are usually stored in a "flat file" table or a relational database management system (RDBMS). Additional hardware and software is therefore required to effectively use GPS for mapping real world features in a suitable fashion for GIS databases. Basic point data collected by receivers must be aggregated to form GIS themes and features, linked to textual descriptive data.

* **When does GPS make data capture in the field more effective than the current digitising process?**

The cost of the map digitising process does not include the cost of making the maps in the first place. In many cases it would most likely be more cost effective to collect new data for a GIS using GPS. Also update information for existing GIS databases would often be collected more cost effectively with GPS. However, we should never forget that other surveying technologies are evolving as rapidly as GIS itself. There may be other positioning technologies that are more appropriate for some GIS applications.

* **Can GPS data be integrated successfully with existing GIS data?**

Where the GIS data has come from maps at scales smaller than 1:20000 there should be no problem integrating differential GPS into existing databases. When accuracies of better than 10m are required for the GPS data, then both the GPS differencing process and the geodetic datum transformations may need to be carefully examined. Integration of data with GIS databases derived from large scale mapping (say 1:1000 or larger) will usually require some expert customisation of the coordinate transformations.

CONCLUSION

GPS and GIS are both technologies which are developing rapidly and being readily adopted for all types of applications. The simple user interfaces to some of these systems disguise the complex processes and fundamental assumptions that are made to collect positional information and set up GIS databases. Many of these issues must be carefully dealt with when linking GPS technology to GIS systems for successful integration to take place. The hardware to enable these processes is generally available "off the shelf" but the appropriate linking software may not always be available. Generally, the technical specifications for GIS and GPS are straight forward enough to permit reasonably simple linking of the technologies for most applications.

It is the real time links between the GPS receiver in the real world and GIS databases that provide some interesting possibilities for future developments. With such technology, field staff responsible for day-to-day management of assets and resources could initialise, update and view spatial data and attributes at any time in the field. Such technology could enable the work practices of many organisations to be optimised very easily and quality control over attributes in databases to be much improved. Such possibilities have already been thought of. For example, the Maritime Services Board is already investigating the use of such technology for managing their Sydney Harbour GIS moorings databases (Colless & Robinson,1993). The directions that the link between GIS and GPS technology in these situations is taking is to develop GPS technology to be an interactive digitising system and a real world equivalent of a mouse on a desktop computer.

REFERENCES

Burrough P., 1986. Principles of Geographical Information Systems for Land Resources Assessment, Oxford, 1986.

Colless R., & Robinson P., 1993. "GIS for Waterways Management", GIS User No.3, May-July.

McElroy S, 1992. "Getting Started with GPS Surveying", GPSCO, Land Information Centre, Bathurst, NSW, 186pp.

Marble D.,F., 1990. "Geographic Information Systems: An Overview", Introductory Readings in Geographic Information Systems, Taylor & Francis, 8-17.

Peuquet D.,J., 1990. "A conceptual Framework and Comparison of Spatial Data Models", Introductory Readings in Geographic Information Systems, Taylor & Francis, 250-285.

Rupert R., 1991. "Travtek: An Advanced Traveler Information System", presented at Institute of Transportation Engineers Annual meeting, Milwaukee, Wisconsin, USA

STDS, 1988. "The Proposed Standard for Digital Cartographic Data", The American Cartographer, Vol. 15, No.1 January 1988.

GIS IN NATURAL RESOURCE MANAGEMENT

Ian P. Musto
National Resource Information Centre
Bureau of Resource Sciences
PO Box E11
Queen Victoria Terrace
Canberra, ACT, 2601

ABSTRACT

The use of geographic information systems and other spatial information technologies in natural resource management differs significantly from applications in the topographic and cadastral database areas. These differences arise not only from the structure of natural resource data but are also from the way they are used to achieve very different information objectives. This paper examines the nature of natural resource data, the issue of uncertainty in data structures and the progress towards common data definitions and continental datasets for natural resource management. The role of GIS in data analysis and information presentation is discussed through examples of recent studies and the use of spatial information technologies in data integration is examined.

INTRODUCTION

Natural resource management includes a wide range of scientific disciplines and applications such as biological/ecological survey and analysis, soil survey, regolith and landform analysis and geological/geoscientific mapping and analysis. Spatial information technologies, particularly geographic information systems (GIS) have been applied to these applications in Australia for some ten years.

The use of GIS in natural resource applications can be considered as a part of a wider spatial information infrastructure, as typified by the work of the coordinating agencies such as State Land Information Council in NSW or the Commonwealth Spatial Data Committee. However, it is important to recognise some fundamental differences in natural resource applications as opposed to the use of GIS in more general land information systems.

After ten years operation, it is possible and useful to characterise the particular issues, problems, opportunities and benefits that are typical of the application of GIS to natural resource management. In particular, it is important to consider three areas: the nature of natural resources data and issues in the definition and storage of entities and attributes; analytical tools and modelling in natural resource management; and the use of GIS as the integration framework for information management to support decision making.

NATURAL RESOURCE DATA ISSUES AND UNCERTAINTY

The principal difference in the application of GIS to natural resource management lies in the nature of the data. In natural resource management we seek to identify, describe and analyse natural phenomena that are diffuse, highly variable in space and time and poorly understood. In contrast most LIS applications deal with an artificial administrative framework developed to satisfy procedural requirements and superimposed on the landscape, for example the Torrens system of land registration.

If we consider the classical spatial paradigm of entities and attributes the difference in the nature of the data becomes clear; LIS data can be unambiguously defined in both dimensions, we can accurately locate the spatial extent of the feature and we can accurately define both the attributes of the class and the attribute value of any member of the class. Any problems that exist in these tasks are not attributable to the nature of the entity; rather, they reflect complexities in the land administration structure or difficulties in isolating and eliminating error.

In contrast, natural phenomena pose difficulties in both dimensions; the nature and shape of the spatial entity may be diffuse or gradational and may change over time. Attributes are subject to error in measurement or description or, more commonly, to a lack of precision in interpretation (identification, classification and estimation). We seek to impose a rudimentary and sometimes simplistic conceptual model on the pattern and process of nature.

Most natural resources data take one of two forms: survey data consisting of accurate and reliable measurements or descriptions of phenomena at sample locations (usually point or small area samples); and what is more correctly called primary level information where sample data have been extrapolated, aggregated or classified to form entities appropriate to the information requirements of an application.

Maps of soil type, vegetation association and geology are all examples of this information where delete spatial entities and attribute values have been interpreted and sometimes interpolated. In the case of a geological map showing both lithology and stratigraphy, the interpretation and classification is extended to three dimensions and includes a prediction of chronology.

In addition to the normal examples of human error: misidentified entities, incorrect spatial locations, transposition of numeric values, typographical error, misspelling, etc., all interpretational processes involve a degree of uncertainty. Measuring and allowing for uncertainty during decision making remains a major issue in GIS use in natural resource management.

For example, consider vegetation association or community data represented as homogeneous polygons in a GIS. In reality, the polygon boundaries are not clear-cut but rather represent a zone of gradual change or ecotone between two classes. In some vegetation types this zone can cover considerable distance. As shown in Figure 1, the actual zone of uncertainty may represent a significant part of the total polygon areas for the vegetation types under consideration. If normal GIS boolean operators are used to combine such data with other data having similar uncertainty as part of a model or analysis, we currently have no clear understanding of the accumulated uncertainty. In these instances the quality of the final information might be assumed rather than real.

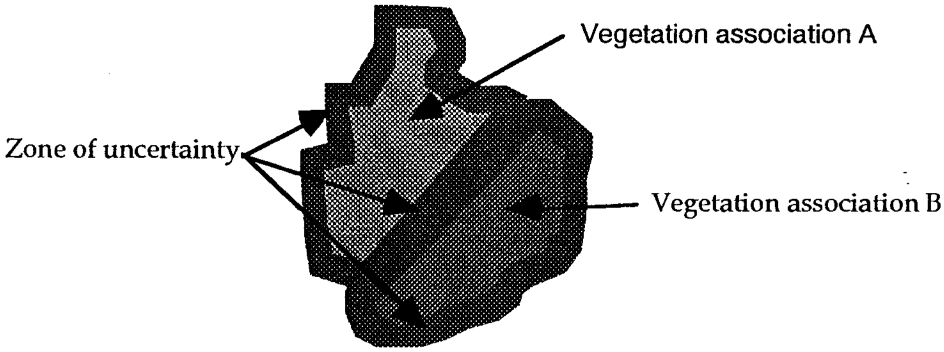


Figure 1. -Zones of uncertainty in 'homogeneous polygon ' information

The importance of accurate survey data, both as the basis for interpretive processes and for post-interpretive correlation is recognised, but Australia has a relative paucity of such data. While most natural resource management agencies are actively involved in 'high-quality' survey, we must continue to utilise transformed information as well as raw measured data and must make allowances for uncertainty.

Uncertainty within any data theme should be considered as an attribute of each entity. 'Confidence estimates' should be stored as part of the database and used as a control in any modelling exercise. For example, we may adopt a rule such as "retrieve only those data where a confidence or precision code exceeds 0.80 for a boolean overlay operation" or may use the precision code as part of a fuzzy logic operation. Recommendations for the storage of data precision information pre-date most GIS installations (Busby, 1979, McDonald, 1984) but are still ignored by some practitioners.

The concise definition of entities and attributes is of particular concern for wide-area natural resource data and information in Australia, ie for datasets that cover the whole of the continent, that cross State borders or cover functional management units such as the Murray-Darling Basin. In recent years a number of programs have been initiated between the relevant Commonwealth and state agencies to address information needs at this scale. Indeed a number of GIS datasets are available, viz: the Atlas of Australian Soils (CSIRO Division of Soils/NRIC) and the Continental Geology (Australian Geological Survey Organisation/NRIC). These datasets suffer from a lack of fine resolution.

Programs such as the National Geoscience Mapping Accord and the National Forest Inventory aim to address the shortcomings of key national datasets. As well as addressing the logistical requirements for the assembly of datasets for the whole of the continent, these programs address the issues of spatial accuracy and precision, data uptake and lineage tracking, common attribute definition and encoding through the use of protocols and agreements. Though this can be time-consuming and tedious, it is necessary to obtain natural resources databases that retain their scientific relevance across administrative zones. It is no longer acceptable that only coarse scale data with poor resolution be used for decision-making at the continental scale.

DATA AND INFORMATION ANALYSIS IN NATURAL RESOURCE MANAGEMENT

Land information systems provide direct benefit from retrieval of data to satisfy an information prescription that is concisely defined by legislation, regulation or operating procedure. While this data retrieval can be complex and involve both spatial and non-spatial criteria, little complex analysis is needed to satisfy most requirements.

In natural resource management, the GIS model is used to provide a common frame of reference within which disparate data sets may be combined to synthesise information not directly contained in any data layer. It is the synthesis of the information that provides the value and benefit to the decision-maker and this invariably requires complex analysis.

Suitability Assessment

Suitability or capability assessment is a technique that is used frequently in natural resource management and one ideally suited to GIS. Here the predictive modelling capability of the system can be applied consistently against management prescriptions. For example, consider the study recently carried out by NRIC for the Department of Primary Industries and Energy to determine a methodology for evaluation of sites for a low-level radioactive waste repository for Australia. The recommended method for site selection is a relatively simple suitability assessment but the use of GIS technology and continental GIS datasets enabled the method to be systematically applied for the whole of the continent in a short time.

Data layers from a number of source scales were combined in a raster GIS analysis at a resolution of 5 km by 5 km. Each cell was classified as 1- suitable, 2 - mainly suitable, 3 - intermediate or indeterminate, 4 - mainly unsuitable or 5 - unsuitable for disposal of low-level radioactive waste.

For each of the data layers used: vegetation, hydrography, hydrogeology, groundwater quality, water balance, relief, soils, regolith, Cainozoic cover, bedrock geology, faults, earthquake hazard, thunderdays, cyclone risk, population density, locations, tenure, road and rail infrastructure, suitability values for every attribute value were determined by an independent panel of experts . A sample of the data layers and suitability values is shown in Table 1.

Theme	Source	Suitability scale
Vegetation	Vegetation of Australia (Carnahan), 1:5000000 scale AUSLIG 1989	1. Suitable: hummock or tufted grass: other herbaceous plants
		2. Mainly suitable: low and medium trees, shrubs; less than 10% foliage cover
		3. Intermediate or indeterminate: low and medium trees , shrubs ; 10-30% foliage cover
		4. Mainly unsuitable: medium trees, 30-70% foliage cover: tall trees, 10-30% foliage cover
		5. Unsuitable: medium trees, greater than 70% foliage cover
Water balance	Rainfall / pan evaporation ratio. 1:2500000 scale Modelled by NRIC using digital elevation model, rainfall and pan evaporation data from CRES, ANU, 1990 1992	1. Suitable: rainfall / evaporation ratio 0-88 (lowest rainfall, highest evaporation)
		2. Mainly suitable: rainfall evaporation ratio 89-125
		3. Intermediate or indeterminate: rainfall / evaporation ratio 126-200
		4. Mainly unsuitable: rainfall / evaporation ratio 201-400
		5. Unsuitable: rainfall / evaporation ratio 401-700 (highest rainfall, lowest evaporation)
Relief	Relief and Landform Map of Australia, 1:5000000 scale, CSIRO, 1969	1. Suitable: plainlands (relief < 30 metres)
		2. Mainly suitable: low relief (30-90 metres)
		3. Intermediate or indeterminate: moderate relief (90-180 metres)
		4. Mainly unsuitable: high relief (180-365 metres)
		5. Unsuitable: very high relief (relief > 365 metres or tidal zones)

Soils	Atlas of Australian Soils, 1:2500000 scale, CSIRO, 1960-68	1. Suitable: very high impermeability; very high water holding capacity
		2. Mainly suitable: high impermeability; high water holding capacity
		3. Intermediate or indeterminate: moderate impermeability; moderate water holding capacity
		4. Mainly unsuitable: permeable; low water holding capacity
		5. Unsuitable: highly permeable, very low water holding capacity; rock
Cainozoic cover	The Geology of Australia, 1:5000000 scale, BMR, 1976	1. Suitable: sand, silt, clay and gravel - alluvial, lacustrine
		2. Mainly suitable: quartz sand - aeolian and residual; minor ferruginous, aluminous and siliceous duricrust
		3. Intermediate or indeterminate: gypsum, halite. clay, sand - evaporitic
		4. Mainly unsuitable: ferruginous, aluminous, siliceous duricrust; calcareous sand and limestone - coastal aeolian
		5. Unsuitable: limestone, terrestrial; minor and sand clay
Bedrock Geology	The Geology of Australia, 1:2500000, BMR, 1976	1. Suitable: fine-grained clayey sediments
		2. Mainly suitable: granitic rocks; gneiss; acid volcanics and intrusives; large mafic intrusives
		3. Intermediate or indeterminate: metasediments; sediments with minor carbonates; coarse-grained sediments; basic volcanics
		4. Mainly unsuitable: sediments with substantial carbonates; ultramafic and mafic intrusives
		5. Unsuitable: limestone, dolomite

Earthquake hazard	Earthquake Hazard Map of Australia, 1:10000000 scale, BMR, 1992	1. Suitable: low frequency, < 0.05 m/s/s 3. Intermediate or indeterminate: intermediate frequency, > 0.05 & < 0.10 m/s/h 5. Unsuitable: High frequency, > 0.10 m/s/s
Population density	Population density by Statistical Local Area, 1:2500000 scale, NRIC, 1992 using ABS 1986 Census Data	1. Suitable: < 1 person per 1000 hectares 2. Mainly suitable: 1 - 5 persons per 1000 hectares 3. Intermediate or indeterminate: 5 - 10 persons per 1000 hectares 4. Mainly unsuitable:10 - 50 persons per 1000 hectares 5. Unsuitable: > 50 persons per 1000 hectares

Table 1. - Sample of data layers used in analysis

To provide an indication of the overall suitability for all locations in Australia, suitability values were combined for all themes using a simple linear addition (areas having the lowest total value being most suitable and areas having the highest value the least suitable).

Besides the total of all themes, a number of other scenarios were investigated. These included restricting the analysis to only those themes having direct impact on radiological safety (the primary factors); and an exclusion method that negated any cell having a mainly unsuitable or unsuitable value for any theme (the worst-case scenario). From these evaluations it has been possible to determine a number of areas for further study using data of higher resolution appropriate to regional and local scale investigation.

This study provides an almost classical demonstration of the application of GIS technology to natural resource management. In summary the important benefits are:

- a consistent sample framework for a number of data sets of different form and scale,
- a consistent, extant method for the assignment of suitability scores to theme attribute values for each theme allowing for critical examination, testing and alteration,
- a simple method for the summation of scores to provide for the evaluation of scenarios and the evaluation of the importance of individual layers within the total model,
- maximum flexibility of approach as opposed to a deterministic pre-classification,

- a dynamic user interface to provide for scenario modelling and overlay of contextual information.

Figure 2. - Gray scale representation of combination of all layers (darkest- least suitable, lightest - most suitable)

Obviously, more advanced techniques for the combination of values could be used and various weightings of layers could be investigated, but these techniques have more relevance at the regional and local scale. At these scales it would be appropriate to include other important data sets such as rare or endangered species and heritage sites.

The utility of the data sets, the modelling method and the user interface can be seen from the fact that the system has been used with different criteria to evaluate preferred landing sites for the Assured Crew Return Vehicle (ACRV) for the space station currently under development by NASA. Here the major criteria were those of both the landing site itself (terrain, vegetation cover) and also trafficability (accessibility, ease of movement) within major zones determined by re-entry position. Further applications of the systems are currently in progress.

Scientific analysis and visualisation

During the early 1980s when GIS technology was under evaluation for natural resource management, one of the chief needs was to move from subjective or objective information and opinion to qualitative or where possible quantitative information. This change is seen as very desirable when considering requirements for the monitoring and assessment of resource management strategies that require a high degree of repetition at high accuracy.

A GIS in itself does not satisfy this objective; it can neither improve the absolute quality of data nor provide high quality information from poor data. However, a GIS can provide a spatial framework for quantitative analysis, modelling and information visualisation. The GIS toolbox can be used to maximise the utility of data and the timeliness of information.

NRIC has been actively developing a comprehensive analysis framework around the GIS that extends the capability in the analysis and visualisation areas. An example of this system capability is shown in work carried out for the Forestry Commission of New South Wales on the South Coast forests.

In this study the GIS was used for two quantitative analyses, viz: a numeric classification of the study area into environmental groups using geological, terrain and climatic variables; and a generalised linear modelling analysis of flora, fauna and other existing site survey data to assess the status of information on rare and endangered species.

In the environmental group analysis, four major environmental gradients were considered: moisture, temperature, nutrients and light. Best available data were assembled in the GIS and the following themes were prepared in a raster database at a resolution of 200 metres:

elevation
relief
roughness (standard deviation of elevation/ mean elevation)
mean slope
maximum slope
minimum slope
nutrient index (P_2O_5 + CaO + K_2O + MgO / Al_2O_3 + SiO_2)
annual mean temperature
minimum temperature of coolest month
maximum temperature of warmest month
annual mean precipitation
precipitation in wettest month
precipitation in driest month
total yearly evaporation
total yearly water balance (precipitation / evaporation)
total yearly radiation.

The elevation and derived themes were processed from a triangulated irregular network (TIN) model of some 6 500 000 points. Climatic variables were processed using elevation data as input to the BIOCLIM (Busby, 1991; Nix, 1986) and ESOCLIM (Hutchinson, 1989) packages. The radiation surfaces from ESOCLIM were corrected for the influence of local slope and aspect (Flemming, 1971).

The 16 primary information layers were classified using a non-hierarchical method based on nearest centroid sorting using the Gower similarity measure (Gower, 1971). The measure of similarity considers all 16 variables for each sample, ie for each cell in the raster. After considering the asymptotes in the error curve, a total of 128 groups was selected.

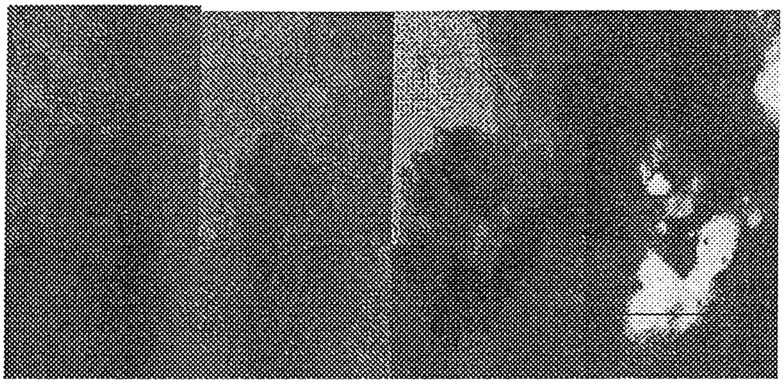
A number of multivariate analysis techniques may be used in the aggregation of groups to form coherent families and to investigate the relationship of the families to environmental gradients. In this study, correspondence analysis was used and the multi-dimensional display facilities of the statistical analysis software were utilised together with the GIS to investigate relationships in the data. The outcome of the ordination was the formation of 16 families which were further characterised as to their principal environmental variables. Colour ramps were used to demonstrate relationships between families.

Site data for the study area represented opportunistic rather than systematic collection and an objective of the study was to evaluate the representativeness of the site data and design an optimum sampling strategy to address the requirement for the prediction of fauna populations. The existing data were analysed as binary presence/absence data using a general linear model approach model in the statistical analysis software. The effect of each of the 16 variables used in the environmental classification was studied using a stepwise method which resulted in the derivation of probabilities of occurrence for the particular species based on the climatic, terrain and substrate values at the known location.

In addition to the uncertainty within each data layer discussed previously, predictive modelling techniques involve a significant level of uncertainty. Information from such analysis is not appropriate as the sole basis for decision-making. In this study, the information is not used in final forest planning, but rather to assess the status of fauna data, to 'flag' areas of significance and to highlight weaknesses in the data used to generate the model.

To highlight information issues, raster models were generated in the GIS for the fitted model and upper and lower 95% confidence limits for selected fauna species. The confidence level surfaces contain information which should be used to qualify the prediction in the fitted surface but in most instances are ignored because of the complexities of visualising the information as a whole. To investigate a better method, one that not only represents the information but also our confidence in the information, the three surfaces were combined in the GIS using the red-green-blue colour model typical of image analysis.

Each surface was scaled in the 8 bit range (0 - 255) and the colour reflected both the fitted value and also the confidence level. For example. a lower confidence value of 255, a fitted value of 255 and an upper confidence value of 255 would be white, ie. high probability with high confidence. A number of scaling factors were investigated for the data to produce a colour range that could be appreciated by users. A grey-shade representation of this information is shown in Figure 3.



a b c d

Figure 3. - Images of a Generalised Linear Model
a - fitted surface, b - lower 95% confidence level surface
c - upper 95% confidence level surface, d - composite surface

The final visualisations were combined with infrastructure data in both planar and pseudo-3D representations to facilitate planning of further survey work and to address requirements for education of local management staff. The information output from the analysis functions appears in the Environmental Impact Statement for forest management of the area.

Although the analytical techniques used in this study have been used in biological classification for some years, their application to systematic evaluation of large areas depends largely on the use of the GIS as the support framework. In addition to providing a consistent spatial sampling base, the GIS provides the initial data analysis tools and the technology to present the final information in an integrated, high quality cartographic form.

Trends

In the lead up to the United Nations Conference on Environment and Development (UNCED) in 1992, there was great emphasis on the concept of ecologically sustainable development (ESD) for natural resource management in Australia. Following UNCED there is particular stress on the need for reliable information on the current status of resources for the whole of the continent and also the trend over time. Thus, data and information analysis in natural resource analysis moves more from the identification of pattern to the identification and measurement of process and the further dimension of time is added to the analytical process.

The use of complex decision support systems which can integrate information from a large number of individual process models will increase as decision-making processes require the integration of biological, ecological, physical and socio-economic factors. Such systems are now available and NRIC has been evaluating the use of decision-support software in conjunction with the GIS to form 20 year predictive models for agricultural production (Veitch et al, 1993).

The roles of the GIS remain; to provide the spatial integration framework, to perform the preliminary data analysis and to provide output and visualisation facilities for the temporal series data as shown in Figure 4. The decision-support software is live-linked to the GIS for both data input and output and further visualisation for multi-dimensional temporal series data can be provided with advanced 3-D visualisation and animation software.

Decision support systems will increasingly become the norm in natural resource management as the process of scenario evaluation becomes an integral part of the policy formulation strategy. The GIS will continue to provide the underlying spatial database, the primary analytical tools and the spatial visualisation framework.

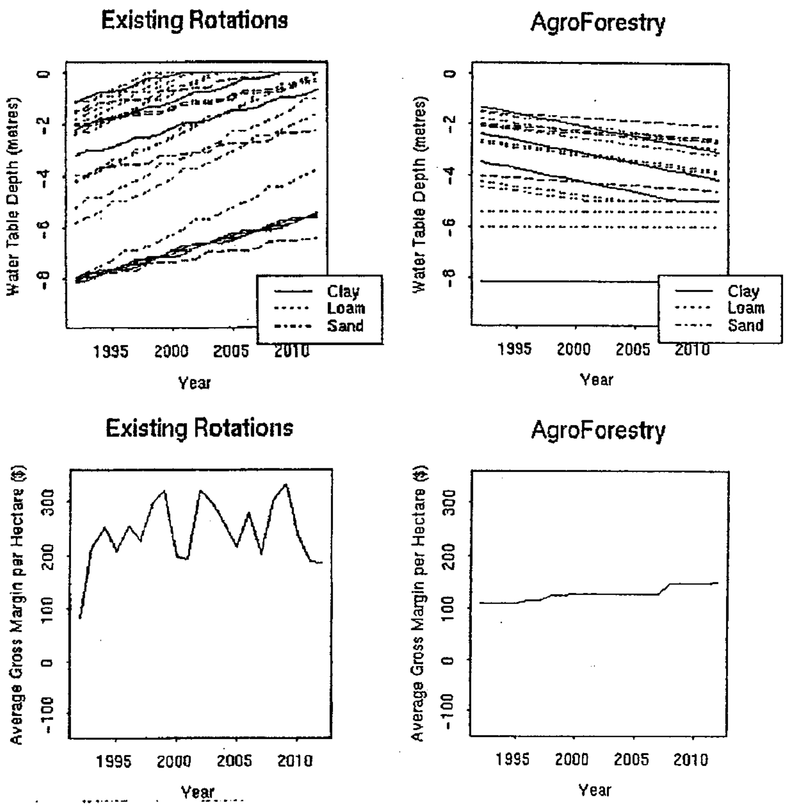


Figure 4. - Existing crop rotation and agro-forestry scenarios for period 1992 - 2012 showing impact on water tables and average gross margin. Produced from WhatIf scenario analysis software

References

Busby, J.R.(1984). Australian Biotaxonomic Information System - Introduction and Data Interchange Standards. Australian Biological Resources Study. Australian Government Publishing Service, Canberra.

Busby, J.R. (1991). "BIOCLIM - A Bioclimatic Analysis and Prediction System". In: Nature Conservation: Cost Effective Surveys and Data Analysis. pp 64-68. CSIRO, Canberra.

Gower , J.C. (1971) "A general coefficient of similarity and some of its properties". Biometrics, 27, pp 857-74.

Hutchinson, M.F. (1989). "A new objective method for spatial interpolation of meteorological variable from irregular networks applied to the estimation of monthly mean solar radiation, temperature, precipitation and windrun". In: CSIRO Division of Water Resources Tech Memo. 89/5, pp 95-105.

McDonald, R.C. R.F. Isbell, J.G. Speight, J Walker and M.S. Hopkins (1984). Australian Soil and Land Survey Field Handbook. Inkata Press, Melbourne.

Nix H.A. (1986). A biogeographic analysis of Australian elapid snakes. In: Atlas of Elapid Snakes of Australia. R. Longmore (Ed), Australian Flora and Fauna Series Number 7. Australian Government Publishing Service, Canberra.

Veitch, S.M., D.P. Fordham and K.W.J. Malafant (1993). "WHATIF? - Scenarios to Explore Land Management Options". Proceedings of Application of Advanced Information Technologies Conference, Spokane, Washington, June 18-19 1993.

LAND COVER MAPPING USING DECISION TREES AND GRASS

Robert Preston and StJohn Kettle
Forest Planning and Environment Branch
Queensland Department of Primary Industries Forest Service
GPO BOX 944, Queensland 4001
Phone 07 234 0124
Fax 07 234 0326

ABSTRACT

The authors of this paper are part of a team developing a Forest Environment and Resources Information System, to be used for land use planning of forest ecosystems. Our work currently involves incorporating remotely sensed data (and other existing data about Queensland's physical environment) into an evolving GIS, and using that data to model vegetation pattern. This paper deals with the GIS and ancillary software which we are using and adapting for our purposes.

The technology we employ may be of interest to others in Australia's GIS community because they are not in wide use: in particular, the GIS GRASS (Geographic Resources Analysis Support System). Being public domain, GRASS is unpromoted by comparison with other GIS packages. This paper is an evaluation-in-progress from a team of users and customisers of GRASS. We will illustrate the discussion by reference to the tasks we perform.

THE CONTEXT IN WHICH WE USE GRASS

FERIS Our section of DPI Forest Service is developing the Forest Environment and Resources Information System (FERIS) as a support for forest and environmental planners in Queensland. The system stores basic information of relevance to the environment, such as terrain, soils, geology, climate and biogeographic region, and more ephemeral information such as satellite images. The system allows for query of this information, and for more advanced analysis of the information, for example the modelling of land-cover or forest structure (Preston, R.A., 1992).

Personnel

The six members of our section's multi-disciplinary team routinely use GRASS (Anon, 1992b). Our experience and skills include botany, forestry, remote sensing and environmental modelling. In computer terms, all of us are UNIX-literate; four of us have programming skills in Fortran, and one of us has programming skills in C, which is the GRASS source code language.

On occasions when the technology gets the better of us, or when a major change to our system is required, we have the support of a computing professional with specific expertise in our computer system and in the installation, customisation and support of GRASS. For example, this professional will soon be installing the recently released version 4.1. He has also designed a display driver specifically tailored to our system.[1]

Technology

GRASS is designed for the UNIX operating system. Our prime platform is a Hewlett Packard 433s UNIX workstation with one high resolution colour monitor and multiple VT220 terminals. This workstation is connected via ethernet with an older Apollo workstation also running GRASS, and with a personal computer running, among other things, the GIS's E-RMS (Anon, 1989a) and MapInfo (Anon, 1992c). The ethernet also links these machines with other computers of Queensland's Department of Primary Industries.

Role of GRASS

GRASS is our main store of raster, and to a lesser extent vector, information. The basic information is passed into GRASS from systems which generate it (e.g. climate and terrain models), or from systems which pre-process it (e.g. image processors), or from other GIS databases.

Our site information is currently stored on our personal computer in a DBASE IV database because of deficiencies we find in the RIM database (Fox, J. 1990) supplied along with GRASS, and discussed below. We pass information from the PC database to GRASS via RIM when we need it, usually for modelling purposes. In future we plan to store site information in a robust and fully relational database interfaced with GRASS.

GRASS is also one of our major tools for working with raster, vector and site information. We use if for display purposes, for generating hard copy, for map amendment (e.g. patching, masking, overwriting), for map query, for map reformatting (e.g. from raster to vector format or vice versa) and, in conjunction with a decision tree package, for modelling.

GRASS has limited capacities for labelling and presenting hard copy vector maps, so we generally pass vector maps to a vector GIS running on our personal computer for labelling and printing. When we need high quality hard copy maps on wide paper we pass them to a cartographic section within DPI, where final formatting and labelling is done before printing to a wide paper colour printer.

---

[1] The Centre for Product and Process Development at the Queensland University of Technology supports GRASS.

USING GRASS: AN OVERVIEW OF PROS AND CONS

Overall GRASS is well-suited to a team like ours, because of our interest in innovative spatial modelling and user-friendly spatial queries. If there is one key feature that attracts us, it is the *flexibility* afforded by access to well-documented source code.

Pros

GRASS is *inexpensive*, being public domain. One can access both the documentation and the source code via Internet, or purchase a US$25 CD-ROM. For some platforms, precompiled source code is available. Having obtained GRASS, we find it efficient to pay for occasional support from a computing professional.

Because GRASS is written in C and designed for UNIX platforms, it is *portable*, and has been ported to PCs running UNIX, and a number of UNIX workstations. There are software houses supporting each port. (Westervelt, 1991 from which much information in this section comes.)

GRASS is *easy to learn and to use.* A novice can begin displaying, querying and reporting on maps within a couple of hours. As the authors of GRASS stress, GRASS has been designed as a programming language offering the user a set of commands, and is fundamentally command-line driven. However, many of the commands can be run interactively. The version (4.1) which has just been released offers an X-Windows interface to these commands.

Because GRASS is public domain, it is comparatively *open* to the experienced user (there are no proprietary tricks to be hidden).

To the user, GRASS feels like an addition to the UNIX community of subroutines. We incorporate GRASS commands, along with other UNIX facilities, into shell scripts, pipelines and so on.

GRASS is *customisable.* The software architecture of GRASS is well-designed, there is a comprehensive Programmer's Manual (Shapiro, 1989), and version 4.0 included 300 subroutines providing interfaces to the data and to the user. Thus for example, a programmer writing a routine which processes the cell values of an input raster map and stores the results in an output raster map has access to libraries of routines for managing the user interface, for reading the cell values of the input raster map, and for writing them out again. This frees the programmer to concentrate on the specific task of the program.

GRASS is therefore an avenue for applications developers to get their algorithms into use. A developer can use the GRASS data format and the GRASS user interface at no cost, instead of either having to design these from the ground up, or tying her/himself to a proprietary interface and data format.

We find the capacity to program within our GIS a great asset. We can implement our own algorithms, and amend those supplied to cope with local requirements. For example, we are increasing the set of GIS's with which we can exchange data efficiently.

GRASS *economises on disc space*, because it automatically compresses all raster maps line-by-line. The greater the repetition in the data, the greater the compression; compression ratios range from virtually 1:1 for spectral images to 100:1 for simple masks.

GRASS can communicate with other GIS's. It is well endowed with routines for im-/exporting vector and raster data from/to other GIS's. These routines have been supplemented by a public domain program written by Vanessa Chewings (Chewings, V. 1991).

GRASS is *growing*, at least in the United States (and the puns quickly pall). Version 4.1 has just been released. There are annual user group meetings in both the US and Europe. The Office of GRASS Integration maintains two electronic mail forums, one for users and the other for programmers, accessible via Internet. There is a quarterly newsletter *GRASS Clippings*. Users in tertiary institutions and elsewhere are developing new code for ultimate integration into the package.

Cons

GRASS version 4.0 has limited capacities for handling vector maps.

The pixel values of GRASS raster maps must be integers, though they may be one, two, three or four bytes in size. The Office of GRASS Integration has on its agenda extension to floating point data and to structures.

Only one attribute (text string) can be associated with each map entity: for example, with each pixel of a raster map, and with each point, line or area of a vector map.

In order to handle site data, GRASS comes with the database RIM and a built-in interface to it. However RIM has severe limitations:

- it supports a rather cumbersome and deficient form of SQL;

- the amendment of the design of a database which has already been loaded with data is technically irksome;

- important types for a site database, like 'date' and 'time' are not supported by the interface GRASS supplies for RIM, although they are supported by RIM itself.

The latest release of GRASS (4.1) incorporates an interface to the relational database INFORMIX.

GRASS is not ideal for cartographic purposes. Though facilities for overlaying legends, grids, text labels and so on are available, they lack the refinement available in some desktop mapping packages, and in more elaborate geographic information systems like ARC/INFO (Anon, 1992a). For example, one cannot vary anything but the color of a line displayed as part of a vector map.

USING GRASS AND DECISION TREES FOR LANDCOVER MODELLING

The way in which we implement our current methodology for land cover modelling (Preston, R.A. and Buck, R.G., 1993) provides a useful illustration of some of GRASS's strengths and weaknesses. Each step of the methodology and its implementation during a recent study is described below.

Step 1: unsupervised classification of satellite imagery. Both the preliminary processing of satellite imagery, and its unsupervised classification, were performed using DISIMP (Anon, 1990). Our unsupervised classifier is AMOEBA (Bryant J., 1990), which we favour because it takes into account both the spectral information about each pixel and the spatial relationships of one pixel to another.

GRASS has a command, *r.in.ascii*, for importing a raster image in ASCII form. This pathway for transferring a image from DISIMP to GRASS was rejected because reading, writing and storing images in ASCII form consumes time and disc space. Instead we made use of some DISIMP image handling sub-routines to write an in-house command, *r.in.disimp*, to import a DISIMP image together with its geo-referencing directly. This task was facilitated by the fact that GRASS separates into separate files the geo-referencing information from the binary array of cell values.

Step 2: the disaggregation of the resulting class map into maximal spatially connected 'clumps' of identically classified pixels. GRASS has a command, *r.clump*, whcih is tailor-made for disaggregating a raster class map into spatially connected clumps of identically classified pixels. This command generates a raster map in which all pixels in the same clump are assigned an integer identifying that clump.

Step 3: the homogenisation of the spectral data by assigning to each clump spectral averages over the clump (with the option of including measures of spectral spread over the clump, such as variance). On a 40 km square scene, there may be of the order of 100,000 clumps. The task of averaging the spectral information over these clumps proved too much for the existing GRASS command *r.average*, which was designed for raster maps with much fewer categories. An examination of the C source code of this command made it clear why it could not cope, and an in-house program, *r.avg_var*, was written to do the same job more efficiently, employing much the same user interface as the original. This program generates homogenised spectral maps in which all pixels in a clump are assigned the same values, namely the spectral averages (or variances, if required) over that clump.

Step 4: the field sampling of a representative selection of bigger clumps for the required mapping feature, say landcover type. Clumps have to be above a certain size to be locatable using a GPS receiver. The task of masking clumps which were too small to for field sampling involved a nice interplay between GRASS and UNIX. Given the map of clumps generated by *r.clump*, the GRASS command *r.volume* generated an ASCII table, each line consisting of a clump identifier and its number of pixels. This table was sorted by clump size, purged of small clumps, and then reformatted to create a text file from which another GRASS command, *r.reclass*, could generate a mask.

Given the original AMOEBA map of classes, and a map of big clumps overlaid with a vector road map, it was possible to develop a clump sampling strategy for a field trip which a) ensured that the sample was stratified by AMOEBA class; b) facilitated the location of clumps which were big enough to locate using GPS technology; and c) facilitated the location of clumps which were close enough to points with vehicular access.

Step 5: using the sampled clumps whose landcover type is known, the construction of a decision tree predicting the landcover of a clump from its homogenised spectral averages. The site information gathered from the field was entered into the RIM database via GRASS's interface *s.db.rim*. The task of associating with each site its homogenised spectral signature was performed by making a raster map whose value is the site identifier at a site, and zero elsewhere; passing the site map together with the homogenised spectral maps to the GRASS cross-tabulating routine *r.stats*; and culling the signatures of each site from the ASCII output of *r.stats*.

We view this procedure as cumbersome, but can see no simpler way to do it within GRASS. The steps above have been incorporated into a shell script.

Step 6: the use of a decision tree to predict landcover clump-by-clump rather than pixel-by-pixel. On the basis of these site signatures together with the landcover classes of the sites, the decision tree algorithm CART (Breiman et al, 1984) generated a tree predicting landcover class from spectral signature.

GRASS has no interface to CART so we have written our own. This involved interfacing a C program delivering the (homogenised) spectral signatures of each pixel to an existing CART Fortran subroutine for classifying them. This task consumed several programming days. The resulting program remains cumbersome and slow. We intend to amend it so that the classification is performed once for each clump, rather than repeated for each pixel in a clump.

Step 7: the use of expert knowledge and other ancillary information to over-write the landcover map where necessary. GRASS is quite well endowed with routines to over-write a map. Among the commands we have found useful are:

- the 'inference engine', *r.infer*, which creates an output map on the basis of a string of pixel-by-pixel inferences about input maps. For example one can apply this inference to two input maps:

if (*inmap1*'s values are in the range 3..10) and (*inmap2* has value 5) then *outmap* has value 5.

- the map calculator, *r.mapcalc*, which creates an output map by applying an arithmetic (or simple logical) expression pixel-by-pixel to input maps. For example:

r.mapcalc outmap='inmap1+inmap2'

adds the two input maps pixel-by-pixel; or more interestingly, for input maps x, a-c:

r.mapcalc *outmap*='if(x,a,b,c)'

sets the pixel value of outmap to that of a if $x > 0$, to that of b if $x = 0$, and to that of c if $x < 0$.

- the map reclassifier *r.reclass*. This is a simple and efficient way to amalgamate (or simply renumber and relabel) the classes of a map.

CONCLUSION AND PLANS FOR THE FUTURE

We intend to go on using GRASS, and are about to install the latest release. As the FERIS database we are developing grows, we anticipate the need to make its information available in electronic form (for example, to regional officies of Queensland's Forest Service). This motivates our desire to improve the portability of GRASS data to desktop mapping packages. Dissatisfied with RIM, we intend to dispense with it and obtain a fully relational database, primarily for handling site information. We would like this database to be compatible with others in use by those who will need access to our information. Because we are keen not to reinvent wheels, we would like to exchange ideas with other users of GRASS with a view to the formation of an Australian GRASS user group.

ACKNOWLEDGEMENTS

Our thanks to the other members of the FERIS team: Rosalie Buck, Graham Lee-Lovick, Dave Jermyn and Doug Ward for their comments.

REFERENCES

Anon (1989), "Environmental Resource Mapping System Version 2.1 User's Manual", National Parks and Wildlife Service of New South Wales, Sydney.

Anon (1990), "DISIMP User Guide", CSIRO Centre for Spatial Information, Canberra.

Anon (1992a), ARC/INFO V6.1 Command References and User Guides, Environmental Research Systems Institute, Redlands, California.

Anon (1992b), "GRASS Version 4.0 Reference Manual", Centre for Product and Process Development (Queensland University of Technology) in association with the Construction Engineering Research Laboratory (US Army), Champaign, Illinois.

Anon (1992c), "MapInfo User's Guide", Mapping Information Systems Corporation, New York.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984), "Classification and Regression Trees", Wadsworth, Monterey, California.

Bryant, Jack (1990), "AMOEBA clustering revisited", Photogrammetric Engineering and Remote Sensing, Vol. 56, No. 1, January, 41-47.

Chewings, V. (1991), "MBMAP: a program to reformat raster files between Arcinfo, GRASS and MicroBrian", Centre for Arid Zones Research, CSIRO Division of Wildlife and Ecology, PO BOX 2111, Alice Springs, NT 0871.

Fox, J. (1990), "RIM Users Manual", University Computing Services, University of Washington, Seattle, USA.

Preston, R.A. and Buck, R.G. (1993), "Improved land cover mapping from Landsat TM by classifying image segments using decision trees and expert knowledge", paper in preparation for the Conference on Advanced Remote Sensing, Sydney, July 1993.

Shapiro, M. (1989), "GRASS 3.0 Programmers Manual", US Army Construction Engineering Research Laboratory, Champaign, Illinois.

Westervelt, James (1991), "An Introduction to GRASS 4.0", Construction Engineering Research Laboratory (US Army), Champaign, Illinois.

Paper presented to the Conference on Land Information Management and Geographic Information Systems, University of New South Wales, Sydney, July 20-22 1993.

WHAT NEXT FOR NATIONAL MAPPING AGENCIES?

David Rhind
Ordnance Survey
Romsey Road
Southampton SO9 4DH
UK

ABSTRACT

The 'template' provided by the topographic map is the basis of virtually all GIS activities. Supply, maintenance and management of this asset is of crucial importance for users. Most such supply has hitherto come from National Mapping Agencies (NMAs). These are common features of nation states but are being profoundly affected by factors both outside and within their control. This paper examines the nature of these factors and how they might affect such agencies. To illustrate both international differences and change through time, the traditional role of Ordnance Survey is contrasted with the contemporary roles of the USA's NMA and of OS. Finally, some speculations are made about the likely changes in the nature of NMAs over the next 20 years.

INTRODUCTION

'Position' is the core of GIS and LIS databases. The description of the geography of individual objects or cartographic features contained in them enables the plotting of maps, the measurement of inter-point distance, the retrieval of data and the process es of data integration and spatial analysis - in short the totality of the activities of GIS and LIS (Maguire *et al* 1991).

The geographical 'template' describing the geometry and place names of a nation is part of the basic infrastructure of a state. Without a coherent, integrated and consistent template, up-dated so as to meet the needs of a multiplicity of customers, land record systems, environmental studies and much else would be impossible. Such a template can now be produced in various different ways. In some countries, satellite remote sensing provides essential detail. The Global Positioning System (GPS) is revolutionising the fixing of control and, in some cases, the definition of detailed points of geography. But the great bulk of the existing templates come from the surveying and mapping activities of National Mapping Agencies (NMAs).

In one form or another, these exist in the majority of the 130 countries of the world with more than about 1 million population. This might be taken to indicate that they are an indispensable part of the infrastructure of the modern state. When the situation is examined in detail, however, huge variations are seen to exist between the roles, organisation, size and financial basis of these different organisations. The British and US federal national mapping organisations, for instance, could scarcely be more different in certain respects - reflecting their nature as parts of two different cultures, arising from dissimilar origins and tailored to the different natures of their national terrain. Many other NMAs are either part of the military or have a close relationship with them.

Is the global pattern of national mapping organisations therefore an accident of history - simply the sum effect of local experiences and decisions, rather than reflecting some basic common need of society across all nations irrespective of their state of development? Is this variegated mosaic also inevitable in the future? The answers to these questions are not simple for the geography of mapping is evolving quite rapidly. This paper sets out to explore the different factors involved, how these affect traditional views of the purpose and role of a national mapping agency, how the British situation fits within the international pattern and what might happen internationally in the future. One assumption made throughout is that digital procedures are certain to become the normal basis for operating, even where this is not currently the case.

FACTORS FORCING CHANGE IN NATIONAL MAPPING

Whatever the historical situation, various factors are now encouraging global convergence and increased homogeneity in mapping. These are:

- rapid changes in survey and mapping technology, diminishing the cost of those tasks which can be done automatically through use of ever-cheaper computers and less skilled labour. The fact that two of the most important of these - satellite-based remote sensing and GPS - are inherently global in scope and execution has its own harmonising effect;

- the growth of internationalism in trading which subsumes mapping as 'just one more service sector' role. This is fostered by the changing attitude of governments world-wide towards their public sector, enforcing more explicit measures of costs and tangible benefits and encouraging the use of commercial contractors, sometimes irrespective of their national origin, for tasks formerly carried out by state employees;

- a growth in awareness by customers of what others obtain by way of mapping service as international comparisons become more commonplace - producing a tendency towards comparable costs of national mapping and use of more similar products;

- a growing need for information in map and related form about the earth as a whole or for large portions of it, rather than for the nation state or fractions of it. The most notable example of such a need is for the set of Global Change scientific programmes now running in many western countries but the effects of the CORINE programme across the EC should not be under-estimated as a catalyst demonstrating the need for more homogeneous environmental data (Mounsey 1991). The need for multi-national digital map data to fuel car guidance systems is another example;

- the moves towards more common standards, manifested by multi-national attempts to harmonise data transfer standards, in use of common digitising specifications in many parts of the world and the *de facto* imposition of standards arising from widespread international use of a few software packages;

Set against these harmonising forces are at least two local factors. The first is the huge cost of obviating differences between existing national systems. Even in areas as well acclimatised to co-operation as France and Belgium, for instance, there are substantial differences in their geodetic datums and hence between map-derived data in the trans-frontier regions. Indeed, within one particular European country, no less than five different levelling datums are in use. Perhaps the most important of the factors slowing harmonisation is the variety of different policies on ownership and distribution of spatial data practised by governments across the world.

In addition to these forces encouraging or discouraging harmonisation, there are also changes in the market for map-type or geographical information. Here the catalyst for the demand for data has been the 'routinisation' provided by the advent of commercial software for handling geographical (or spatial) data. The growth of Geographical (Coppock and Rhind 1991) and of Land Information Systems (Dale 1991) has been well-chronicled. Though most applications of these to date have been record-keeping (rather than analytical) ones, it is clear that the making of traditional maps is not the justification of most investments. The scope for linking together different data sets, initially collected quite separately, gives users of GIS something extra for minimal cost: in principle, the more combinations of data sets exist, the more applications can be tackled.

What we can deduce from all this is that huge external forces are affecting mapping (and much else). National mapping agencies have almost no control over these forces. Yet, since national mapping agencies have traditionally been the suppliers of the basic topographic framework (or template) on which others (e.g. soil surveys) have produced their maps, these forces may well in turn have a huge effect upon what other organisations can do and, subsequently, on what the public can obtain. For instance, a recent claim is that NASA is the only agency capable of creating a consistent global Digital Elevation Model at XY resolutions of around 50m and much higher Z resolution. If this were to come to pass, it would affect virtually every national mapping agency in the world, especially those who already market such a product. Whether this is a good or a bad thing depends on many other considerations - some of which we will explore later.

UNRESOLVED PROBLEMS IN MAPPING

It was emphasised above that a substantial number of factors affecting their future role and success were outside the control of national mapping agencies. There are however also many key problems they face which *are* within their ability to solve. Few, if any, national mapping agencies yet seem to have satisfactory technical, logistic or economic solutions to them yet this will affect their long-term success. To illustrate the importance and complexity of the issues, a number of them will now be outlined.

Multiple representations

A map or a computer representation of it is a model of the 'real world', excluding those elements of the latter believed irrelevant to the purposes of the model's creator. The included elements are also simplified in a way which does not invalidate the initial purpose of data collection. Thus the model may be said to be at some 'level of abstraction'. Traditional cartographic generalisation creates change in the level of abstraction.

All of this has severely practical consequences for the GIS user. The reasons for it are very simple. Almost all contemporary GIS applications are based upon use of a topographic base map or 'template' to which other data sets are fitted. It follows that any inconsistencies in the templates used to create different data sets will generate mis-matches between those data sets. In non-expert hands, such mis-matches could cause literally fatal errors since current generations of GIS do not operate with any significant knowledge of the data they contain and many operations and applications no longer necessitate graphic expression.

Ignoring algorithmic or user-induced error, mis-matches between topographic templates come about by three means. The first is through use of two or more templates which exist at different levels of abstraction when collecting other data which are subsequently overlaid, one on the other. Overlaying data compiled on 1:10,000 scale OS mapping with other data compiled on OS 1:50,000 scale mapping may be unwise. Though OS experiments show that this problem is generally smaller and more insidious than the author had expected, this conclusion can not be generalised to apply to all such 'within family' comparisons. The second cause of mis-matches is the use of two topographic templates at similar levels of abstraction but which have been compiled quite independently and thus differ significantly in the detail of their geometry. Combined use of data whose geometry was controlled by two topographic templates compiled to different specifications and assembled by different procedures - and thus having no familial relationships - is almost certainly a recipe for disaster. This is typical of the situation where data referenced to OS 1:250,000 and, say, Bartholomew's 1:200,000 digital data are overlaid. The third - near criminally negligent - cause is where these effects are maximised through use of both, say, OS 1:50,000 and Bartholomew's 1:200,000 scale topographic template data when collecting other data sets which are then overlaid.

There is then demonstrable scientific merit and - if an appropriate accounting metric could be applied - economic benefit from use of one family of templates, especially if the most 'real world-like' products could be used to produce the 'less correct' ones. For national mapping agencies, this need for a consistent template is a huge bonus - in effect only OS or its NMA equivalents can normally provide cohesive national coverage of detailed and high quality data from which other consistently related products can be spun off. But the need for within-family consistency and for algorithmic generation of cartographically acceptable siblings from their parents poses great technical and operational difficulties. However great the difficulties, these will need to be faced in the next few years, at least in customer-funded NMAs.

Up-dates or 'snap shots'

The increasing use of photographic products as base maps in some parts of the world is totally understandable. Where no detailed up-to-date line mapping exists, this is almost certainly cheaper and faster than traditional survey methods. Users can obtain orthophoto products which match many traditional survey accuracy specifications - provided that adequate geodetic control exists and information invisible from above but required as annotation (e.g. street names) is available.

Somewhat to the author's surprise, OS has recently had experience which suggests that the consequences of such an approach may turn out ultimately to be highly unpopular with certain users. The class of users concerned is that which records their own information on the topographic template and where its position in relation to particular mapped objects is important. Examples of this include utility companies recording buried plant in relation to the edge of a pavement or scientists recording soil sampling points in relation to the edge of a wood. Where the detailed co-ordinates of the 'base object' are subsequently changed in the course of up-dating, this causes significant work on the part of the user if s/he wishes to preserve the previously established relationship. In the case of OS, where mapping is continually up-dated, such co-ordinate shifts are consequences of progressive improvements in survey accuracy and can be expected to diminish in number and significance over time. Where photographically based 'snap-shots' are used, however, each 'snap shot' will produce slightly different co-ordinates for an object from those obtained previously.

This highlights a latent difference in philosophy between different NMAs. Many of those still struggling to create a topographic template in computer form have understandably given little attention to up-dates, regarding these as a matter for the longer term. The OS' approach, at least so far as the basic scales data are concerned, has been to maintain a record of what customers already hold so that, if an up-date is to be purchased, only what has changed since the original purchase can (in theory) be supplied (see Coote and Rackham 1992). The analogy is that, as a tyre of one's car wears out, a new one can be bought and fitted - rather than having to purchase a whole new car. This has some implications for the user's operations and even more for his or her software. In principle, however, it seems a logical way to proceed with those customers who are major data holders.

Thus the national topographic database turns out to be a curious entity when seen from the view point both of the data custodian and of the user. In the first instance, there is typically not one but multiple databases, each with different properties yet supposedly representing the same 'real world'. In no sense is this analogous to statistical summaries of a detailed database such as of bank accounts. Adherence to use solely of topographic templates from a NMA will minimise the problem but not obviate it. In the second instance, the core database may be maintained as a series of largely unrelated national 'snap-shots', with the entities being map sheets, or as a fully functioning seamless database in which 'real world' objects are up-dated in a 'normal IT database' sense by transaction processing. The financial attractions of the data collection mechanism for the first of these options, at least in some parts of the world, may lead to databases which cause considerable user expense and frustration 'down-stream'. In the OS experience at least, consistency in the record of unchanging objects is highly valued by users. We expect this view to become more widespread as users' migrate from pilot-based operations and establish their own substantial databases. In practice, then, mechanisms for obtaining consistently positioned objects from remotely sensed imagery and differentiating these from new features or ones which truly have changed position are going to be essential requirements in future.

Datum and projection differences

OS has abandoned many traditional surveying practices in favour of GPS-based procedures. The manifestation of this is that nearly 5,000 of its 6,000 trigonometrical stations will be discarded for survey purposes (OS 1992a). The clear advantages of GPS to NMAs and to users render the greater use of GPS inevitable. Yet, as Ashkenazi (1993) and Dodson and Basker (1992) have shown, the different principles on which position is now fixed render the old and the new incompatible in certain respects, especially in the Z values. Moreover, adoption of a European-wide datum consistent with use of GPS and accompanied by a standard European projection system will inevitably produce poorer quality mapping in a country at the edge of the continent.

Suitably transformed digital map data can already be produced by the Survey which shows sub-metre level differences when compared with GPS co-ordinates. Such wholesale conversion of their databases is unattractive to many users of detailed topographic mapping. But OS recognises the problem that 'lack of fit' between the existing mapping and GPS may cause for those users who intend to employ both (likely to be the majority of GPS users). We have taken action to alert users to this matter (OS 1992b) and are actively seeking their opinions on it (which will inevitably not be unanimous). Though this is all caused by 'technology-push', its most important consequences will be financial, organisational and legal ones.

Protection of intellectual property

In the next major section we deal with the policy issue of how to fund a national mapping agency, as seen by different governments. Little of this is within the control of the individual NMA. But, for all those mapping agencies which are expected to recover any of their costs, it becomes ever more important to preserve the Intellectual Property Rights inherent in the created artefact. Action on this is within the scope of the individual NMAs.

Piracy of information is now occurring on an increasing scale: for instance, a court case is currently on the point of being brought in Germany in regard to use by publishers of mapping which was pirated from another country's NMA by an organisation in Eastern Europe. OS itself has had experience of theft of its material. In this regard, OS and those other NMAs which do not follow the US federal model (see below) have the same interests as all commercial enterprises. The protection of our intellectual property may force NMAs into the same tactics as commercial enterprises to 'finger print' their products viz. create deliberate but insignificant errors within the data sets. This poses a moral problem for NMAs and also a severe technical one given the desirability of maintaining the seamless topographic database in the way described earlier.

The pan-national dimension

As 'internationalism' and, in particular, 'Europeanisation' become ever more real, what will be the effects on NMAs? The need for global map coverage is manifested by the appearance of the *Digital Chart of the World*, derived from military maps at 1:1 million scale. It is also clear that there is some civilian need for pan-European topographic databases at about the level of abstraction typical of 1:250,000 or even 1:100,000 scale mapping: the EC's Eurostat has compiled an interim data set of European statistical area boundaries and various commercial agencies have compiled digital maps for route planning purposes. Small (e.g. 1:1 million and smaller) scale maps are best regarded as caricatures of reality: the immense problems of assembling data sets collected from different sources and relating these to such low resolution and highly generalised mapping are already well-known (Mounsey 1991). For this reason, the more expert amongst the user community are coming to appreciate the limitations of such smaller scale mapping, especially when used in a GIS, and are seeking more detailed and accurate topographic templates. As the higher resolution products are approached, however, the consequences of inconsistent datums, map specifications and compilation procedures becomes ever more clear and the amount of work to harmonise the whole multi-national mosaic becomes ever greater.

The Comité Européen des Responsables de la Cartographie Officielle (or CERCO) is a pan-national group of NMAs extending over most of Europe. Currently it has some 30 members. Whilst its primary function is to exchange information on matters of mutual interest and concern, CERCO was approached by Eurostat in 1991 to assemble a Europe-wide topographic template. In the event, this proved to be impossible in the very short time scales that Eurostat required but the Institut für Angewandte Geodäsie in Germany is assembling a set of administrative boundaries from data available for individual nations. Their investigation of the availability of such data and the form in which it exists demonstrated huge differences between countries. Perhaps of greater long term significance, a group of countries - Britain, France and the Nordic group - provided staff for a CERCO Technical Group based in Paris whose primary role was to examine the desirable scope and character of a Europe-wide topographic database. The group has reported successfully and a group of those NMAs who wish to participate on a commercial basis is now proceeding under the banner of the MEGRIN programme, outside of the non-executive CERCO. Inevitably, however, such commercially sensitive collaboration is complicated by variations in the terms of reference of different NMAs, differences in the level of development of their own databases, etc.

In the longer term, it is inconceivable that we will not have a European civilian database of topography equivalent to at least 1:100,000 scale mapping, possibly to 1:50,000 and perhaps even to 1:25,000. At present, we are far removed from that point because of differences in civilian specifications and much effort will be required to overcome these problems. Institutional and political factors are just as significant for NMAs as the technical ones noted earlier. In many respects, the future of European NMAs is a microcosm of the future form of the European Community. If this becomes a much more close-knit community, the prospect of one European Mapping Agency may even come about. At the time of writing, this seems an unlikely prospect but the unpredictability of history - demonstrated graphically by the events in the former Communist bloc in the period 1989-91 - renders it impossible to rule out. It is also conceivable that the very changes in technology and remote sensing discussed earlier may facilitate the creation of new pan-European alliances for particular tasks e.g. to make a photomap of Europe from satellite imagery or pan-national consistent Digital Elevation Models. In some such cases, it may even be easier to start again than to attempt to convert existing material into a consistent whole, thereby by-passing the NMAs and their existing materials. Clearly all this represents a major challenge for NMAs.

THE ROLE OF NATIONAL MAPPING AGENCIES

Before we delve into the future, it is sensible to summarise the traditional Anglo-Saxon view of a NMA and describe its current British and US federal government variants.

The Smith model

Smith (1979) considered the needs of government and society on a 'from first principles' basis. From this, he deduced the need for particular categories of information. His arguments are set out in detail in Rhind (1991); here they will merely be summarised. He argued that government had two primary responsibilities. These are to ensure order and to encourage material and other progress (such as encouraging economic growth, raising health standards and conserving the environment). The first is a mandatory responsibility and the second is optional.

Ensuring order involves the state in safeguarding national security and protecting individual rights to life, property and liberty. The price paid for this by the citizenry is the surrender of some individual rights and resources in the common interest. Smith argued that security purposes demanded three criteria be met by the topographical information collected: that it is up-to-date, is homogeneous (i.e. to a common reference system and common specification) and is continuous in spatial extent (i.e. exhausts the national territory).

So far as encouraging progress is concerned, Smith advanced three reasons why government "..should in any case be involved in the collection of topographic information". These he summarised as:

- when it is impractical for individuals or particular sections of society to provide it for themselves [and, by implication, when they need (c.f. desire) it];

- when it is less expensive for the nation as a whole if it is provided centrally (permitting economies of scale);

- when the pooling of many needs for a variety of information makes it cheaper to be met this way than to be met separately (permitting economies of aggregation).

The present author has argued (Rhind 1991) that some of these *could* be achieved without the government itself actually carrying out all of the operations but that, on the other hand, there are other good reasons why the state must be intimately involved in mapping through a NMA.

The first erosion of the 'state as provider of free information' principle came so far as OS was concerned at least as early as 1966. That principle was progressively eroded but formed the basis of much of the Survey's operations until 1979. Since then, change has rapidly occurred towards a more market economy approach.

The UK government's approach

The UK government's policy is that subsidies are undesirable and must be discontinued at the earliest possible moment. Central government funding of the maintenance of even part of the basic scale mapping is seen as a subsidy. Thus, unless customers can be found for goods or services, these should not be produced. Mapping and geographical information is therefore being viewed as a commodity. The concept of 'government' being treated as a single customer for any commodity is explicitly ruled out (see the Chancellor of the Duchy of Lancaster's speech in Hansard, November 4th 1992). In principle therefore OS should be seeking payment from all those individual branches of government which need mapping as contributions to meet the *total* cost of these activities.

Several consequences follow from adopting this general approach (which is not specific to OS but the Survey is more affected by it are than many other UK government agencies). These have been set out at length in Rhind (1992b) and include:

- the effect that 'ability and willingness to pay' will certainly have on the re-mapping of rural areas where fewer customers exist than in more urban areas;

- the shortening of time scales for returns on investments;

- the growing reliance on revenue streams to continue projects and for investment for the future;

- the huge reliance on assessment of market needs - as expressed by the market itself. What gets shown on OS maps will increasingly have to be viewed as equivalent to the components and accessories of a motor car; Superplan is the first step along the road to customised products. Given the emphasis on market needs, the continuity of the historical record is also much more difficult to guarantee.

That said, it should be noted that government does not expect a return on sums expended in the past for *ab initio* survey or 'one-off' digitising i.e. these are treated as sunk costs. In addition, the cost of maintaining the core topographic mapping of the country (1:1250 to 1:50,000 scale mapping) is still subsidised by the taxpayer, at least in the short term.

Finally, one other aspect of OS' work is little-known. This relates to its role as the provider of official advice. Hence the author, by virtue of his office, is the advisor to the Secretary of State for the Environment on all surveying and mapping matters. This is reflected in OS staff sitting on certain government committees but the role has hitherto been exercised in a less dramatic fashion than is commonplace in mainland Europe. There, NMAs are not simply NMAs: they are (sometimes explicitly in name) National Geographical Institutes which provide a wide variety of official information and advice. Placing a value on such a role in monetary terms is not easy but is accepted as of considerable value by many other governments.

The US federal government approach

National topographic mapping in the USA is provided by the National Mapping Division (NMD) of the US Geological Survey. This works at a basic scale of 1:24,000 and complete paper map coverage of all the USA but Alaska is available at this resolution though only about 10% of it is yet in computer form. Complete recent topographic map coverage of the entire country is available at 1:100,000 scale in both paper or digital form. The size of the country - some 30 times the size of Britain - puts this coverage in perspective (though GB is mapped at basic scales on average at 100 times the US basic scales resolution). USGS have been highly innovative in areas such as orthophoto mapping. In many respects, however, they resemble the pre-1979 Ordnance Survey in that marketing of their products, especially to non-government audiences, is not an overwhelming priority.

The most obvious difference between the UK and the USA is that most already-existing federal government data is disseminated to any one at the cost of reproduction and dissemination (see, for instance, Rhind 1992a). Moreover, the federal government relinquishes all rights over that data. This has been claimed by many (e.g. Maffini 1990) to have formed the basis of the growth of US dominance in the GIS market place. Sadly, the situation is much more complex than that: dominance began long before much US data was available. The role of military funding of pilot production, the size of the home market and sheer entrepreneurial skills in the commercial sector seem to have been at least as significant.

The distinction between the approaches of the UK and USA governments leads to huge differences in price of products: though at present there are few truly comparable products, Wingham and Wolf (Pers. Comm. September 1992) have shown that European prices for certain Digital Elevation Models are highly variable and at least one or two orders of magnitude more expensive than US ones (the British prices being about half way up the world list). It is inevitable that such price differentials have *some* effect upon the market, especially as the prices of hardware and software fall in real terms. Set against this, however, is that OS users in Britain did not rate price level - but rather price stability - as their main worry in a recent Round Table exercise carried out by the Association for Geographic Information. This may reflect different views by different sectors of the market place.

Though the difference between the UK and the USA is apparently stark in both policy and price terms, it is simplistic to see the British NMA as commercially-driven and the USGS as driven by public good considerations. In 1991, some $30M (or 18.5%) of the $162M budget of the National Mapping Division was derived from the Reimbursable Program and from sales of goods. In most cases, this Program takes the form of a 50/50 cost share with other agencies to pay for map revision, digitisation, etc. Those who enter into such agreements are generally anxious to have data quickly for certain purposes or areas or in a particular form. About 60% of the Program is with other federal agencies but 8% of it is with individual States, counties and municipalities and the remainder with other, non-federal sources. Steps are being taken to increase the relative size of the Program. Thus it seems clear that USGS/NMD's priorities are influenced by customers who are willing and able to contribute half of the total costs of an operation.

Such statistics demonstrate that comparisons of the two systems are much more complex than is commonly supposed. It is, for instance, important to note that the principle of supplying data at cost of reproduction does not apply amongst all states and lower levels of government: there the situation is much more varied, some having opted for explicit cost-recovery practices of one type or another. The May 1993 Geo-Data Policy Forum in Washington DC demonstrated a multiplicity of views over the role of government in collecting and disseminating data. Nevertheless, it is clear that the NMA data available in the USA is largely what federal government of one form or another requires for its own purposes. In that sense, non-government customers simply take what they get (though they get a great deal), having relatively little direct influence on the product specification or even on what is collected. In Britain, by government definition, all customers are equal - if they can pay.

WHAT THE ORDNANCE SURVEY IS COMMITTED TO DOING

Most users care little about policies or principles: their concern is with product availability, specification and price. It is therefore appropriate now - by way of example - to summarise the current situation in Britain. Here the national mapping agency obviously works within government's overall policies. Based on these, the policies and actions which OS will be following over the next five years have been defined. Though these have been set out in full in Rhind (1992b) and were delivered to the November 1992 conference of the AGI, it is appropriate to summarise them here. They assume continuing user interest in products, manifested by acceptable revenue streams, and continuation of existing government policies. The policies and actions are:

- completion of digital mapping coverage of Great Britain (to a higher specification than is current) by 1995 rather than the date of 2000 set in 1990;

- further improvement and re-engineering of the 'basic scales' data into the National Topographic database in the period up to 1997;

- greater concentration of resources on mapping of change in urban areas;

- greater predictability of up-date in *all* areas;

- continuing introduction of new products (such as Address-Point, the national address database) and new services to meet market needs, the former at least largely arising from 'spin offs' from the National Topographic Database. This will provide greater choice for customers;

- achieving full cost recovery by 1997, as defined in Rhind (1992b). This pre-supposes that some contractual solution to the mapping of commercially unprofitable areas is agreed;

- collaboration with other organisations wherever mutually beneficial to create or market new products;

- price stability within the constraints set by market size and government policy. The objective is to achieve higher cost recovery levels through greater volume of sales and cost reductions rather than price rises wherever possible. Since OS is not a profit-maximising body, any reductions in cost of operations will be passed onto customers once government cost recovery targets are met and future investment is safeguarded;

- withdrawal of individual product lines (or parts of them) which make a loss unless costs are met by a sponsor.

Apart from the dramatic speed-up in completion of the digitising of the 230,000 basic scale maps of Britain, perhaps the two most significant features of this summary are the publication of detailed targets and the concentration on innovation of new products and services.

THE NEXT 20 YEARS IN MAPPING

Many current investments in GIS are designed to have a pay-back in as little as three years though for some the period is as long as a decade. Given the rapid decrease in real terms of computing costs, the wise manager therefore considers when best to make a proposal to invest. The data needed looms increasingly large in such a consideration. What then is the prospect for data providers in general but national mapping agencies in particular over the next twenty years? The first observation to be made is that any such prediction will differ somewhat between individual NMAs because of their different starting positions. The second observation is that any prediction will be wrong in some respects. The third is that it is much easier to forecast the near future than the distant one. Given that, we start with what Rhind (1992b) has already set out as the OS expectation in the period up to about 1998:

- much wider use of GIS-type capabilities will exist, especially by non-traditional users, though these facilities will be regarded as no more novel than is word processing now;

- partly as a consequence, the use of multiple data sets in combination will also be routine, with these data sets being held by different organisations;

- computer networking for access to OS (and other) data will be widespread and routine - as it already is in France for many people through the *Minitel* system in their home. Indeed, the indications are that 'reaching out for information' will be no more technically difficult than is the current use of a telephone. It should not be imagined that all this will be restricted to the Developed World: the rapidly diminishing costs of facilities ensures that many Developing Countries can leap-frog other, supposedly more technically sophisticated, nations. In Mexico, for instance, the census-taking organisation runs a much more sophisticated satellite-based computer networking system than their equivalent in Britain, all run by young staff trained in the USA. It is no coincidence that Mexico has embarked on plans for a national GIS;

- there will be some form of a National Land and Property Information System (or Service) which operates as a gateway and bridging facility to a multiplicity of data sets and analytical tools;

- some customers will wish to have different types of information to those presently collected (such as the heights of properties and possibly the appearance of them to facilitate three dimensional visualisation);

- even for those customers who still want paper maps - and there will be many - printing instantaneously on demand from the database is likely to be commonplace. 'Map sheets' may well, for the professional market at least, become a meaningless concept - they will have maps covering only the area they need for any given task;

- historical information may be routinely regarded as part of the remit of the national mapping agency and be supplied by the organisation on demand through the same distribution channels as contemporary data;

- there will be both much greater international collaboration *and* competition to supply geographical information such as the topographic template.

The longer term is even less certain but, fortunately, many newly operational tools and even electronic consumer items seem to have a gestation period of around a decade from their first appearance in research laboratories to widespread consumer acceptance. Assuming this will continue, we can also assume that the shape of the next 20 years is being assembled now. Here then is a personal set of scenarios which might come to fruition by the early part of the 21st century:

- national mapping agencies will be much more customer-focused but many different organs of the state will become customers, not least because use of some NMA products will be specified in legislation as necessarily used for certain purposes;

- this apparent return to the 1960s will not be a consequence of renewed state corporatism but rather because the costs of having multiple different representations even at about the same level of abstraction will have been quantified - probably following some gross error in GIS operations caused by incompatible data having become embarrassingly public;

- the form of national mapping agencies will be very different to that now. All will be smaller and cheaper to run and will contract out some functions;

- at least some NMAs may no longer exist, having gone the way of certain national airlines. The most aggressive NMAs will offer services world-wide, in conjunction with a constellation of private contractors;

- truly national land and property databases (or multi-purpose cadastres) will be in place in some countries which have a comparative advantage (such as Britain?) and in the process of being introduced elsewhere. Such a transition will offer consultancy opportunities to those organisations with demonstrably successful experience in dealing with 'large scale' mapping, especially if this is used as the core of the land and property database;

- a much larger fraction of topographic data will be collected remotely by various sensors, some operating outside the visible part of the electromagnetic spectrum. Information extraction techniques will be used to derive as much as possible from the images. On-ground survey will only be used where essential - such as for land use (c.f. land cover), postal addresses, property boundaries and rights of way;

- the international dichotomy between ways of regarding geographical information - as a commodity or as a state-provided benefit funded by general taxation - will be different. It is however completely impossible to predict the out-turn at present.

CONCLUSIONS

The future roles of individual NMAs will inevitably remain largely controlled by policies of their governments but some convergence can be expected as a consequence of technological change and changes in customer awareness and sophistication. Rarely, if ever, however can the uncertainties and challenges facing NMAs have been greater. As a result, many are likely to change significantly in coming years. Ordnance Survey will certainly be a different organisation in many respects in a decade. My intention is that OS will continue to play a key role in Britain and, subject to finance and political approval where necessary, play an increasingly large role in any international market for mapping and geographical information.

ACKNOWLEDGEMENTS

Thanks are due to Steve Guptill and Neil Smith who commented on the first draft of this paper. Responsibility for any errors of fact or interpretation are the author's alone.

REFERENCES

Ashkenazi V. (1993) GPS as an integrated positioning system. *Proceedings, Survey and Mapping Alliance Conference, Paper 8E*, Keele.

Coote A.M. and Rackham L.J. (1992) Handling up-date: issues in spatial data maintenance. *Paper 2.30 Proceedings of the Association for Geographic Information Conference*, November 1992, Birmingham.

Coppock J.T. and Rhind D.W. (1991) The history of GIS. In : Maguire D, Goodchild M F, Rhind D W (Eds.) *Geographical Information Systems: Principles and Applications*, Vol. 1, 21-43, Longman, London

Dale P.F. (1991) Land Information Systems. In : Maguire D, Goodchild M F, Rhind D W (Eds.) *Geographical Information Systems: Principles and Applications*, Vol. 2, 85-99, Longman, London

Dodson A.H. and Basker G.A. (1992) GPS: GIS problems solved? *Paper 1.12 Proceedings of the Association for Geographic Information Conference*, November 1992, Birmingham.

Maffini G (1990) The role of public domain databases in the growth and development of GIS. *Mapping Awareness* 4, 1, 49-54.

Maguire D, Goodchild M F and Rhind D W (1991) (Eds.) *Geographical Information Systems: Principles and Applications*, Longman, London

Mounsey H.M. (1991) The creation and use of multi-source, multi-national-environmental GIS: lessons learned from CORINE. In : Maguire D, Goodchild M F and Rhind D W (Eds.) *Geographical Information Systems: Principles and Applications, Vol.* 2, 185-200, Longman, London

OS (1992a) Ordnance Survey Triangulation Pillars - what is to become of them?. *OS Information Paper 8/1992.*

OS (1992b) The impact of the Global Positioning System on national mapping policy. *OS Consultative Paper 2.*

Rhind D.W. (1991) The role of the Ordnance Survey of Great Britain. *Cartographic Journal* 28, 188-99.

Rhind D.W. (1992a) Data access, charging and copyright and their implications for geographical information systems. *International Journal of GIS* 6, 1, 13-30.

Rhind D.W. (1992b) Policy on the supply and availability of Ordnance Survey information over the next five years. *Paper 1.22, Proceedings of the Association for Geographic Information Conference*, November 1992, Birmingham (also in *Mapping Awareness*, 7, 1 and 7, 2, 1993).

Smith W.P. (1979) National mapping: a case for government responsibility. *Conference of Commonwealth Surveyors 1979, Paper M2, Cambridge*

DEFINING THE QUALITY OF SPATIAL DATA: A DISCUSSION DOCUMENT

N S Smith and D W Rhind
Ordnance Survey
Romsey Road
Southampton SO9 4DH, UK

ABSTRACT

Defining the quality of spatial data is a non-trivial exercise for it is a multi-dimensional problem confounded by user perceptions. Yet it becomes ever more necessary as such data becomes more widely employed in GIS and LIS which are operated by users who understand little about abstruse computer procedures or survey jargon. Moreover, failure to provide coherent and understandable definitions not only complicates the use of data but may open the supplier to litigation. This paper proposes some basic descriptors of quality and goes on to suggest how these might be operationalised.

INTRODUCTION

A selling point of contemporary GIS and LIS software is their ease-of-use. As the numbers of users of such systems continue to grow, the average level of understanding of what underpins the systems can be expected to reduce still further. In the short term at least, the complexities of individual algorithms and variations in quality of the data used will become meaningless and invisible to most users of 'black box' tools.

This is a serious matter. Rarely is there only one answer possible from use of a GIS, at least when it is used for analytical purposes. Results from such systems depend on embedded algorithms, the quality of code produced by the system creator, the integrity of the data and the skills of the user. Creators of products have an obligation - certainly moral and possibly legal - to define and publicise the nature and reliability of their products. Nowhere is this more clear than in the case of spatial data provided by respected national, state-owned organisations such as National Mapping Agencies: Rhind (1993) has argued that the 'topographic template' they produce underlies and thereby controls most of the subsequent data sets used in all GIS and LIS. This paper therefore discusses the issue of quality in spatial data. Its purpose is to foster discussion.

If the increase in accessibility and use of geographical data is one 'trigger' for the creation of quality measures, another is the change in the nature of the data creation process. Map-making has long been a craft-based activity, with a long apprenticeship, rigid rules and intensive checking procedures typical of series mapping. The need for greater cost-effectiveness, and the trend to contracting out work, plus the advent of Total Quality concepts and the consequential reliance upon individuals to monitor their own work to a much greater extent has changed the data compilation norms. The new norms demand formal statements of the processes by which a product has been forged and validatable descriptions of its quality.

CONCEPTUAL ASPECTS OF QUALITY

The very nature of quality is an elusive concept. In principle, the enlightened consumer or the lawyer may regard quality as performance against product specification. In practice, a domestic product such as a light bulb is judged by most consumers to be of high quality on a variety of innate perceptions relating to its ability to perform effectively, reliably and efficiently in energy terms. Few consumers have the ability to test whether the product meets detailed technical specifications - and which may not, in any case, be published: statistics such as 'mean time between failures' may be highly confidential.

Where a failure of a domestic product occurs, the manufacturer often solves the problem by replacing the faulty goods on the basis that the failure is 'one off'. The implication is that a failure of quality control or an act of God rendered defective one good out of many identical ones. In a database environment, however, where few customers will ever receive supposedly identical products, any gross errors revealed will call into question the entire integrity of the database. The obvious solution is to monitor the process of creation as well as assessing individual end-results. It also suggests the need for independent audit, a point to which we return later.

Despite these caveats, quality is most meaningfully defined as a measure of performance against specification. How that assessment is done will depend on the nature of the specification and, in particular, the individual user's requirements: it may well be that a product is entirely 'fit for purpose' for one user even if it falls short of the published specification.

Two other general matters must be considered here. The first is that the user is only concerned with *product* specification and quality. It is entirely conceivable that the database from which the product is 'spun off' will have very different and superior quality characteristics. The relationship between the two sets of characteristics is primarily a marketing consideration. The second matter is that the national legal framework under which the transfer of data takes place is an important factor: there is reasonable evidence that data given away or sold cheaply will in some countries attract much less legal liability than data sold for a high price. Moreover, whether the legal treatment of data occurs under *caveat emptor* ('user beware') or under consumer legislation which assumes the supplier is to blame for failure unless proven otherwise will also be an important consideration: the normal risk assessment carried out prior to marketing new products must take this into account. It follows from all this that product specifications and quality statements become a necessary part of marketing of products.

MEASURING THE QUALITY OF PRODUCTS

We can distinguish two categories of information which the user may require. The first is product-specific: it includes details of what the product is and the quality assurance checks that have been applied. It may also include, in general terms, the means by which the product came about. The second comprises details relevant to the data for the area supplied e.g. how many features of a given type are included? It is obvious that the first *may* continue to be relevant if the data is sub-divided and passed on to another user: even if the first user generalises the data and extracts one portion of the area, these characteristics would still be relevant to the quality statement of the new data. The second category is however invalidated in whole or in part by any post-initial supply operations.

There are at present two different approaches to measuring the quality of products. These are:

- measuring quantitatively the adherence to certain previously defined product specifications, such as the US National Mapping Standards (see, for instance, Thompson and Rosenfeld 1971, Merchant 1987);

- defining the processes by which data has been produced and the Quality Control and Quality Assessment procedures in place. This forms the basis of such quality standards as the British Standard BS5750/International Standards Organisation ISO9000 standards (to which OS is now in the throes of certification). It is also the basis of much American standards work preparatory to the production of the Spatial Data Transfer Standard (now FIPS 173), being described as 'truth in labelling' (Grady 1990).

FACTORS TO BE TAKEN INTO CONSIDERATION IN QUALITY STATEMENTS

This section is based upon work commissioned by Ordnance Survey, experience with the UK National Joint Utilities' Group specifications, other documents such as the Australian Data User Guide for 1:250 000 scale digital mapping (AUSLIG 1992) and the US Spatial Data Transfer Standard (SDTS: see USGS undated) plus papers in the academic literature.

In different ways, the AUSLIG Data User Guide and SDTS are pioneering documents. Having studied them, however, we believe they under-represent the complexities of the issues and the difficulties users will have to face. For those reasons, we believe there is a need to extend the concepts and measures described therein and to explore the many complications which arise in making quality statements. On this basis, we now deal in sequence with the various categories of data specification and related quality tests. The categories are not exclusive and, indeed, necessarily interrelate. The majority of these categorisation refer to the characteristics of the product but one relates to the process by which it has been created.

Content

This indicates *what* is represented in the product - in other words, which features or objects that exist on the ground and have been selected by the data creator are represented in the map or data. Classifications of reality are, of course, many and varied (Land 1991, Tosta 1993). As one example, European NMAs do not use a consistent classification (though one exists for 1:25,000 scale mapping, it is not used by any one NMA). Furthermore, as user needs change, the content of databases may also be expected to evolve.

More fundamentally, the classification to be used raises major strategic and tactical questions for all national organisations: how should they cope with different needs of different users (demonstrated by the different definitions of wetlands by different agencies in the USA - see MSC 1993)? To what extent should NMAs operate on a Highest Common Factor approach i.e. should they be expected to record and codify features which are solely region-specific? Viewed with 20/20 hindsight, the emergence over extended periods of different national and regional typologies of the natural and built environment parallels the earlier emergence of biological typologies. Seen from an international perspective, however, the mapping community has less commonality than do even the biologists.

More prosaically, the simplest specification of the content of data products is provided by the list of codes used to describe entities. Together with the completeness category and the depiction category (see below), plus a glossary of terms, this forms the semantic specifications of the products. In practice, the content specification will be manifested both by qualitative and quantitative 'facts' arranged in lists. For example, this might entail a simple statement that the product will show Bench Mark positions (by means of a co-ordinated point with related text). In some cases, the user will wish to have more synoptic information, such as how many Bench Marks might be expected in such an area. The constitution of the objects or entities should be indicated - for example that all buildings should contain a building seed point (if that is the specification).

Given such specifications and sampling of the data for quality checking, it is possible to define quality measures which show, by means of a statistical summary table for the data domain chosen, how well the content elements are represented (e.g. how many types of feature are present on the ground and in what frequency?) and the level of adherence to the specification: what, for instance, is the proportion of buildings that do not have seed points?

Temporal truth

This represents the *time element* of the content specification. In other words, how long after the moment of creation of one example of a 'feature' included in the specification can a customer expect to see the event represented? And how long after a feature has disappeared will it still remain in the 'current' database? This is in part a matter of product policy but customers should be entitled to know when changes in reality will be reflected in the mapping. The specification is likely to be mostly quantitative - e.g. "90% of features of a particular type in a particular sort of area will be updated within x months". Some products, such as OS' ED-LINE (comprising boundaries of the 100,000+ Enumeration Districts used for the 1991 Population Census) will, of their very nature, never be updated: the specification must say so. Honesty is essential - if a particular area or a particular type of feature is not going to be updated for thirty years, this must be pointed out in the specification. In short, the specification should be set to match expected performance and user needs. Failure to achieve such performance is a management issue.

There is, of course, the possibility that an example of a type of feature defined in the specification may *never* appear on a map or in the database whilst other examples elsewhere are included. This context-sensitive generalisation is commonplace in manually produced mapping of 1:50,000 scale and smaller. Because of it, small scale mapping provides an incomplete but simpler and hence 'readable' model of the world. The mechanism by which to summarise it quantitatively is, as yet, far from clear (see, for example, Buttenfield and McMaster 1991). What we *do* know (see João *et al* forthcoming) is that automatically generalised mapping from larger scale mapping or more detailed databases gives generally more predictable and hence certifiable mapping and data products than does its manual counterpart. This suggests that both definitions of information sources and the generalisation parameters applied to produce the product may become necessary parts of the specification, perhaps as part of the lineage category (see below).

The quality measure will be how well the data creator or custodian matches the specification with its up-date service.

Depiction

This covers *how* objects are represented. More formally, it describes how real world constructs are generalised into data objects - whether digital or analogue. Thus in the OS Land-Line.93 specification, buildings are represented as lines and logically unrelated 'seeds' or label points.

Quality measures will include the accuracy of depiction: for instance, are all buildings represented by the appropriate codes as defined in the content list? What is the proportion of place name spellings that are correct (compared with a standard - and how should this standard be defined and sourced)? Where the level of abstraction is such that symbolism rather than miniaturised form is used, what proportion of entities is correctly described? For example, road widths and carriageway natures are often classified into a few categories in 1:50,000 and smaller scale mapping: an accuracy measure might therefore be what proportion of roads is shown in the correct form (e.g. dual or single carriageway, >4m rather than <4m)? All of this pre-supposes that 'truth' is known. Since knowledge of complete truth would enable all errors to be fixed, it also pre-supposes that truth is known only from sampling studies and that the sampling strategy is such as to permit inferences about the population of errors as a whole.

Geometric fidelity

This determines how closely the position of discrete objects shown on the map or in the database must agree with their position on the ground. The specification must set limits of tolerance. This element has three main components:

- *how the shape of an object compares with reality.* Strictly speaking, simple measures of this type are unable to differentiate between error and deliberate 'feature-shift generalisation'. Moreover, any relative or absolute accuracy statement must ideally also take different aspects of shape into account;

- *how accurately the features are represented in relation to each other*. This defines how accurately it should be possible to measure inter-point distances in the data or on the map. It is probably best represented by a systematic error plus a factor proportional to distance - e.g. 0.5 metre ± 50 parts per million);

- *how accurately the position of a feature or object is recorded relative to the reference framework* e.g. the British National Grid. This specifies the absolute accuracy of a single point relative to the whole co-ordinate system.

Operationalising these is far from easy: shape differences are not easily defined in ways meaningful to the user and the sample is inherently biased because measurements are only possible from and to well-defined points (such as road junctions). It may also be that these definitions should be disaggregated by feature or object type (e.g. buildings as compared with contours) and by area. The user is likely to be less interested in the overall accuracy of a product than in his/her area of interest but geographical disaggregation of quality values could well be problematical as many organisations do not necessarily know the detailed genealogy of any one sheet. Finally, there may be some contents of the database for which conventional geometric accuracy statistics are meaningless, such as any feature computed from another (e.g. an 'area seed', derived by algorithm from the boundaries of a building or land parcel).

These specifications must be set so that they are realistic, meaningful and facilitate an honest evaluation of the quality. It is probably better, for example to state that "70% of all well-defined points on products derived from 1:1250 maps can be expected to be represented within 0.5m of their true position with respect to the National Grid" than "the RMSE of 1:1250 mapping is 0.352m" (these figures are examples only).

Topology and connectivity

These measures specify how well features relate together in spatial terms, independently of a metric. Hence features that intersect in reality must do so in the data. Similarly, if a road is always to the east of a canal, it must be represented that way. If the data contains explicit relationship information, for example links and nodes or polygon coding, then the extent and accuracy of this must be specified. Specifications will be both quantitative and qualitative (such as that connectivity in a particular product will be considered only in two dimensions: this has the consequence that roads crossing at bridges will be shown as nodes). Users might reasonably expect to see specifications such as "100% of all boundary polygons will be closed; at least 99% of all road junctions will have explicit turn information..." Other elements of topology that could be included are in matters such as overshoots and undershoots in "cleaned" data, polygons having finite area, etc. Edge-match could sensibly be presented as a special case of a connectivity measure.

Again, the quality measure is how well the marketed product achieves these specifications.

Logical consistency

Mapping products have a defined form - whether data or graphic. In data products, the customer is able to get a specification for the files from documents (in the OS case, these include British Standard 7567 (NTF) or DXF manuals). In graphic products, there may be either a defined specification (if a printed product) or a range of specifications (if plotted on customer demand). Logical consistency, as defined here, is the specification for the physical manifestation of the product.

Data validation is used by OS to test whether the files to be despatched to customers conform to the specification. It seems intrinsically reasonable that customers could expect a statement that a file has (or has not) been so tested. Clearly when an external file organisation is used, testing of the internal definition must be carried out. Any divergence from the standards must be explained. There are some analogies in graphic products - for example is the printing registration correct? Is the grid correctly positioned? Is the sheet name relevant? Does the cover match the sheet? In this category at least, the aim must be to achieve 100% compliance. Failure to achieve this in data products would normally render null the purpose of supplying the file or paper map.

Lineage

In many standards, lineage is seen as being an important indicator of quality. However, whilst *a priori* knowledge might suggest one supplier of data or process to be more reliable than another, this is an inadequate basis on which to judge quality in any measurable way. In principle, the source of the data set is irrelevant provided it meets the specifications set. Some definitions of lineage (notably SDTS or FIPS 173) also require great detail of data creation processes to be given, such as those to ensure no loss of precision on transformation. In our view this is a production specification, not a product specification.

As indicated above, many organisations do not necessarily know the genealogy of any one map sheet or storage unit. Moreover, there is a trend in at least some NMAs towards continuous up-date in which a number of different processes may have been applied in different places to different classes of object at different times. The consequence of this is that genealogical information must become object- and area-specific if it is to be of any real value. It follows that maintaining such information must become increasingly onerous with the passage of time and up-date of the data.

Documentation

Implicit in all of the above is that products must be documented and that this documentation itself forms part of the quality of the product. A feature of much early documentation from NMAs world-wide (including OS) is that it was written from a producer viewpoint, stressed survey techniques and physical transfer characteristics and was written by surveyors or data experts. Measurement of the quality of documentation is non-trivial, especially given the needs of disparate users.

THE FORM OF QUALITY STATEMENTS

Quality statements are for the benefit of the users. This means that they must be:

- *Relevant.* There is no point in publishing specifications or quality statements if they do not directly relate to the customers' needs for information. Bland statements that tell the user nothing do not support the reputation of the data supplier;

- *Quantitative* wherever possible;

- *Accessible.* There is a need for a structured set of documents ranging from free leaflets to chargeable reports - possibly even commissioned accuracy testing work. This implies additional expense for many data-providing agencies. Information about quality has itself got value and limits of free or chargeable access must be set;

- *Easily understood.* There is a danger of confusing and alienating users if there is too much detail. However it can not be *assumed* that users will not want the detail. There is already evidence that detailed quality information is required in data supply contracts in Britain at least;

- *Attributable.* The source of information and the techniques used to obtain statistics must be known and stated if appropriate. This will mean that there should be a single authority within the data supplier for supplying statistics and quality information. Such an authority is in any case required for BS 5750/ISO 9000 certification;

In short, the quality statement subsumes the product documentation. There is a requirement not only for structured documentation that specifies the characteristics of the products, but also documentation which defines a related set of procedures to test that the products truly do have those characteristics and which reports on the results of applying the tests.

CONCLUSIONS

No-one doubts that specifying quality of products is of increasing importance to users and producers alike. The problem is how to do it in such a way that will meet multiple objectives, be meaningful to users, relatively economical to produce, can be widely implemented across many organisations and yet not be violated by the first action of those users in loading the data into their GIS or LIS.

It will be evident from all that has gone before that producing meaningful quality statements is non-trivial. This can be illustrated by considering the characteristics of Land-Line.93, a product now being launched by OS which is derived from 1:1250 and 1:2500 scale mapping covering the non-mountainous areas of Britain. In one sense this is a relatively simple case since little generalisation of the data is involved. However both the mapping and the database were created by a number of different processes over an extended period of time (up to 40 years for the source mapping). As a consequence, it is impossible to identify the actual quality of almost any single point. On the other hand, recent processing of the data has been strictly controlled and some aspects of the content, depiction, logical consistency and the topology may be defined quantitatively. In certain areas where specific accuracy tests have been carried out, the relative and absolute accuracy can be defined. Since these tests have not been carried out everywhere, however, the worst case has to be inferred. For all these difficulties, Land-Line.93 is one of the best data sets of its kind: specifying quality of other data sets will certainly be more difficult.

Some other key issues for discussion arising from this paper are:

- how can we simplify the descriptions suggested in earlier parts of the paper and in other documents such as the SDTS manual? Complexity may reflect the nature of our data and its life history but is inimical to use of quality standards by the vast majority of users;

- can we devise quality standards which are in such a standard, quantified form that GIS systems can operate on data automatically taking into account the quality characteristics of these data? Without this, all judgements on data use will have to be taken by humans;

- in practice, how will users really be able to judge if the data not only match the original specification but are also suitable for a specific task? There seems to be little or no experience of this at present;

- how will organisations be able to concatenate data statements, probably of variable quality, from all previous organisations involved in handling the data? Does this mean that quality statements will grow in length and complexity as the data's life proceeds?

- will organisations become liable for errors or omissions in quality statements pertaining to data which they inherited or obtained from another organisation and then enhance and distribute? If so, this may well engender collection of new data sets with certifiable characteristics rather than re-use of old ones;

- just as some algorithms or code in software are commercially valuable and treated as secret, some characteristics of data are likely also to be of commercial value. Thus OS has the capability to transform Land-Line data onto the equivalent of WGS84 datum: this transformation will be accurate to sub-metre levels and hence is much more accurate than transformations presently built into GPS receivers and commercial post-processing software but the numerous parameters may well remain commercially secret. Quality statements of the kind described in SDTS and earlier in this paper pre-suppose all details of the data are made available in the public domain;

- how will we solve the problem that the (significant) costs of creating quality statements falls upon data creators but in general the benefits arise elsewhere? Will data creating organisations with stretched budgets devote part of them to devising documents which bring them no financial benefit? Will the creation of data quality statements therefore require legal or financial incentives?

- how will we make international comparisons of data set quality where there is no international agreement on the form these should take and where some organisations do not produce quality statements?

- who is responsible for assessing the quality of the quality statements? In principle, this can not be the data provider. Perhaps there needs to be a class of individuals who perform the GIS equivalent of financial auditing - the data quality auditor.

The mapping/GIS/LIS community is at an early stage in the definition of quality statements for geographical databases. We hope that this paper fosters discussion of this internationally important topic.

ACKNOWLEDGEMENTS

Thanks are due to Nick Chrisman and various of our colleagues in Ordnance Survey for their helpful suggestions. Responsibility for any errors of fact rest with the authors.

REFERENCES

AUSLIG (1992) *Topo-250K Data user guide.* Australian Surveying and Land Information Group, Belconnen.

Buttenfield B and McMaster R (1991) (eds.) *Map generalisation - making rules for knowledge representation.* Longman, London

Goodchild M G and Gopal S (1989) (eds.) *Accuracy of spatial databases.* Taylor and Francis, London

Grady R K (1990) The lineage of data in Land and Geographic Information Systems. *URISA Jl.* 2, 2, 2-6.

João E, Herbert G, Rhind D W, Openshaw S and Raper J F (forthcoming) Towards a generalisation machine to minimise generalisation effects. In Mather P (ed.) *Geographical Information Handling*, Wiley, Chichester.

Land N (1991) *The classification and coding of spatial data.* Unpublished CNAA Ph.D thesis, University (formerly Polytechnic) of East London.

Merchant D.C. (1987) Spatial accuracy specifications for large scale topographic maps. *Photogrammetric Engineering and Remote Sensing* 53, 7, 958-61.

MSC (1993) *Towards a Coordinated Spatial Data Infrastructure. for the Nation.* Mapping Sciences Committee, National Research Council, Washington DC.

Rhind D W (1993) What next for National Mapping Agencies? *Proc. LIM/GIS conference*, Sydney, July 1993

Thompson M.M. and Rosenfeld G H (1971) On map accuracy specifications. *Survey and Mapping* 31, 1, 57-64.

Tosta N (1993) A sense of space. *Geo-Info Systems* 3, 5, 21-5.

USGS (undated) *Spatial Data Transfer Standard.* US Geological Survey, National mapping Division, Washington DC.

ERROR HANDLING IN GEOGRAPHICAL INFORMATION SYSTEMS

Alan Rogers
Research Officer
Department of Minerals and Energy
Western Australia
100 Plain Street
EAST PERTH 6004
Australia

ABSTRACT

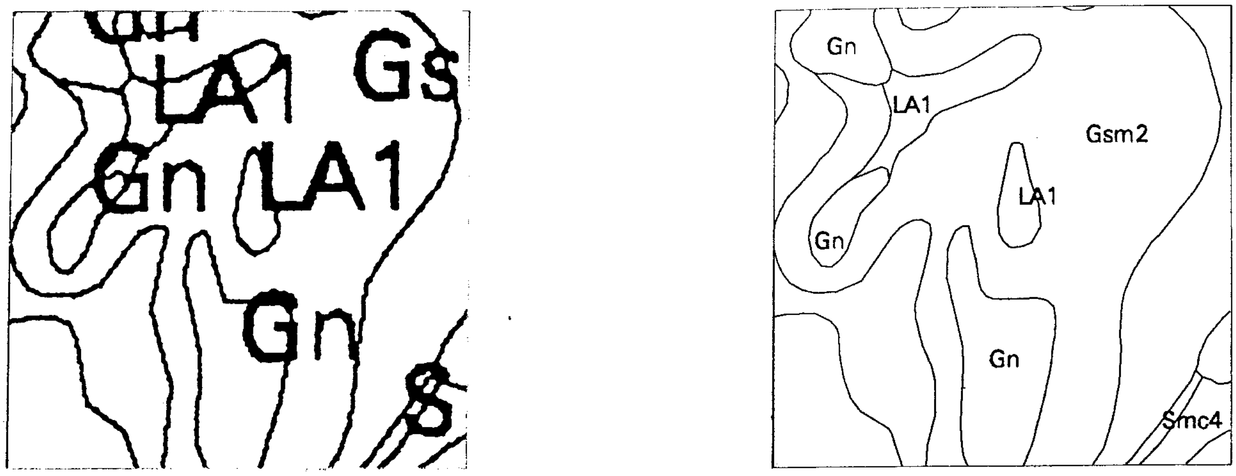
Errors that occur in spatial databases are numerous and varied. Anomalies that Cartographers have, over the centuries, handled in the production of hard copy maps take on an entirely new dimension in a seamless, scaleless computer database. Some errors, such as those associated with projection are no different in the digital world, but when we digitise the simplified or generalised line of the hard copy Cartographer we find that our digital line has coordinates significant to sixteen decimal places and appears to be of extreme accuracy! Although this is not an error in the true sense of the word, anyone can use the data at any scale and may be completely unaware of the validity of the data at that scale. The hard copy map, on the other hand , has a fixed scale and the data should be valid at that scale. This paper examines these errors associated with digitising and map scale and discusses the method in which the Department of Mineral and Energy handles then when analysing G.I.S. data.

INTRODUCTION

Errors are not a new phenomena to both Cartography nor Geographical Information Systems (GIS). Since the first maps were created Cartographers have been using various techniques, such as projections to minimise error, or have deliberately introduced error to make the map easier to use.

Before going any further perhaps the term "error" should be defined in the context used in this paper. Chrisman (1991) defines error as the "deviation (or distance) between a measurement and true value." The position of a feature on a map will almost never reflect the exact ground position. The distance between the two is what will be referred to in this paper as error.

In hard copy maps these errors are usually not of great concern to the user. The scale and projection of the map are (or should be) well displayed. There is some danger if the map has been photo-enlarged and the original scale is not shown, but the more course appearance usually alerts the user that the map has been altered. This is not the case with computer assisted mapping, there is no fixed scale and there is usually no facility for displaying the current scale and the projection used. A data set, for example, that was digitised from a 1:50 000 map is fine for viewing at 1:50 000 or a smaller scale, but a user who is constantly zooming in and out and panning around usually has no idea of the current scale of the display. The line symbology stays spider web thin as if it were digitised at that scale. (see fig 1)



(a) *(b)*

Fig 1. The difference between an enlarged map and the digital map - (a) plotted at 1:50000 and enlarged to 1:12500 (b) plotted at 1:12500.

This paper examines the effect of errors on the Cowaramup Environment Geology GIS Project, recently completed by the Department of Minerals and Energy, Western Australia. It goes on to explain a simple method of displaying errors on-screen, so that allowances can be made during analysis of data.

COWARAMUP ENVIRONMENTAL GEOLOGY G.I.S. PROJECT

Background

The Cowaramup Environmental Geology GIS Project was based on the 1:50 000 Environmental Geology map series produced by the Department of Minerals and Energy, Western Australia. This map series was produced using conventional cartographic methods, such as scribing, masking, proofing and plate making. One hard copy map was produced showing geology, geomorphology and basic raw materials and well as a basic framework of roads, contours, hydrographic and cultural features. These maps are concentrated around urban areas and the South West of the State and are intended to be used as tools for planning authorities and for the management of basic raw materials in these areas.

It was proposed that the Cowaramup sheet which was the next map in the series be produced as a Geographical Information System and be used as a pilot project for subsequent maps in the series. The GIS was developed using Arc/Info software on an IBM RISC System 6000 workstation. Arc/Info's TIN module was also used for surface analysis.

Perceived advantages of the G.I.S.

Apart from the obvious benefits of using computers for mapping, such as zooming in and out, easier updating of data and output at any scale there are other real tangible advantages that a GIS has over conventional mapping of this type.

The range of data can be significantly increased. Other related data sets were be utilised that were not able to be included using the analogue techniques. Mining and Land Tenure, bore information, town planning zones, conservation areas and soils are examples of extra data sets. Remotely Sensed data was also used to add a further dimension to the project. More information means that better quality planning decisions can be made.

An extensive textual database exists for most data sets, particularly geological sample sites, bores and tenements. Users can view on screen much more information that can possibly be portrayed on a conventional map. A new geological survey system was introduced so that data collected was directly compatible with Arc/Info's database. It is hoped that in subsequent geological surveys of this kind a data logger and a Global Positional System (GPS) unit will be used to further enhance and standardise the survey procedure. This will cut the data entry time out altogether therefore saving a large percentage of data collection time.

Maps can be customised to users demands. If someone requires a map only showing particular data or subsets of that data, GIS provides the tools to do this. The data base can be queried to pick out only the relevant data and symbolise it in a meaningful way. Data can also be exported to other specialist systems for enhanced analysis or production, such as three dimensional modelling or map publishing.

GIS provides tools for analysis of data sets. One data theme can be overlaid and intersected with other themes to achieve completely new themes. Data sets can also be derived using surface modelling techniques such as slope and aspect.

Cowaramup 1:50 000

The study area is Cowaramup 1:50 000 1930-3 extending westerly into Mentelle 1:50 000 1830-2 to include a narrow strip of coastline. The map is part of the State Medium Scale series, using Universal Transverse Mercator Projection on the Australian Map Grid (see Fig 2.).

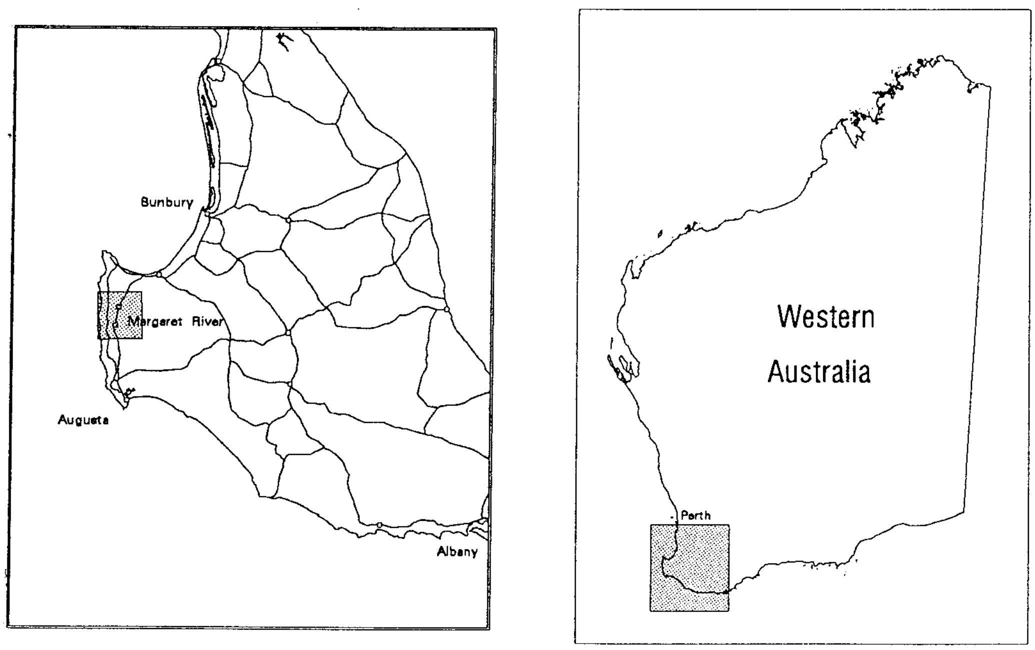


Fig 2. Locality map.

The region is renowned for it's world class surfing beaches and a producer of quality wines as well as being a popular tourist area. The major activity in the region is agriculture and the principal town is Margaret River. The map spans two Local Government Authorities, Shire of Augusta-Margaret River and the Shire Of Bussleton.

Data sets

The following data sets were included in the project.

DATA	SOURCE
SPOT panchromatic image	Dept Land Administration
Contours	Dept Land Administration
Bore hole data	Dept Minerals & Energy Water Authority WA
Slopes	Derived (TIN) (Air Photo)
Geology	Dept Minerals & Energy
Mining Tenure	Dept Minerals & Energy
Soils	WA Dept Agriculture
Cadastre	Dept Land Administration
Town Planning Zones	Local Govt Authority, Dept Planning Urban Development Dept Cons. & Land Management Dept Minerals & Energy
Geological Sample Sites	Dept Minerals & Energy
CALM Estates	Dept Cons. & Land Management
Leeuwin - Naturaliste Management Areas	Dept Cons. & Land Management
Roads	Dept Land Administration
Control Points	Dept Land Administration
EPA Areas	Dept Minerals & Energy
LGA boundaries	Dept Land Administration
Faults	Dept Minerals & Energy
Coastline	Dept Minerals & Energy

Table 1. *List of data sets used for the Cowaramup Environmental Geology GIS.*

TYPES OF ERRORS

As mentioned earlier errors encountered in Geographic Information Systems are numerous and varied. Burrough (1986) lists fourteen possible sources of error classified into three groups - "Obvious sources of error, Errors resulting from natural variation or from original measurements and Errors arising through processing". The first two groups of errors can be regarded as normal "cartographic" errors and are often encountered in the traditional methods of map making. The third group of errors are unique to the computer and often can only be corrected by the computer scientist when designing systems.

Often the data sets required for a GIS are held by another organisations where the sources of error are unknown. Quite often very little or no supporting information, such as date and scale of capture, method of capture, coordinate system used or density of observation is supplied. This makes it very difficult to place any credence on the results from analytical observations. As can be seen from Table 1 the Cowaramup GIS project required much of it's data to come from external organisations. None of that data was supplied with supporting Meta data and it was only through further research the relevant error details were determined. The importance of this "Meta-Data" cannot be over stated if a GIS is to be of any real use. This will be discussed in greater detail later in the paper.

Although there are many sources of error in GIS including operator's mistakes, this paper only addresses error associated with digitising and map scale. The reason for this is that these errors probably constitute the major source of error and are the easiest to control.

Digitising

Because it is impossible to represent the curved and undulating surface of the earth on a flat, uniform plane, therefore the boundaries on a map cannot be regarded as absolute. Every point, line or polygon or pixel has some degree of error in it's position. If this error factor was known or able to be deduced then a confidence interval can be ascertained that can be taken into account when analysing that data.

The conversion of points or lines on a map to digital form through the use of digitisers is the most common way of capturing data, although we may see the use of scanners play a bigger part in the future. Presuming the map we are digitising is free of distortion due to paper shrinkage or warping, there are numerous factors that can affect the accuracy of digitising.

The precision of the digitisers can affect the accuracy of capture. Consider digitising a 1:50 000 map. If the digitiser has a precision of 0.025 millimetres the error introduced will be 1.25 metres on the ground. Many cheaper digitisers have a precision of 0.25 millimetres which produces an error of 12.5 metres. Digitisers can also introduce small errors due to the way they are constructed, although these are being minimised with modern production techniques (Milford 1993).

Most maps are produced with an inherent positional accuracy of 0.1 millimetres. When digitising a map at 1:50 000 this would translate to 5 metres on the ground. This is the best accuracy you can achieve from this map (Milford 1993).

Experienced digitisers operators usually maintain an accuracy of about 0.25 millimetres (Milford 1993). This can escalate if using inexperienced staff. This will be discussed later in the paper.

The digitising of curved lines on a map also presents a problem. Computers do not store these as curved lines, but a series of straight lines to represent the curve. If these chords are too far apart then the line appears less like a curve and it introduces error in the middle of the chord. The smaller these lines are together the better (and more accurate) the curve is represented, but the storage requirement of the line is greatly increased.

Map Scale

The scale of the source material is an important factor in determining the accuracy of the data. The thickness of a line on a map translated to ground units represents the greatest accuracy you can achieve using that map. For example: a 0.25 millimetre line on a 1:50 000 map translates to a ground value of 12.5 metres. The same line on a 1:250 000 map would yield a ground value of 62.5 metres. Burrough (1986) calls this the "area of uncertainty".

Thematic or continuous data, where the data has been classified into distinct units presents us with another type of boundary discrepancy. In the real world, the boundaries between naturally occurring phenomena is seldom clear. Geologists have to make expert decisions determining the "cut off" point between two geological units. The boundary may be quite clear such as a rock outcrop, but when confronted with one type of sand gradually changing to another type of sand over fifty metres or so, the boundary can be said to have an accuracy of plus or minus 25 metres. Soil scientists, botanists, demographers or anyone classifying (particularly natural features) data are forced into drawing a boundary between the two and calling one side 100% type A and the other side 100% type B. While this appears to be a problem with the type of resource being mapped, the density of observations used in classifying would be valid at a particular scale. Viewing this data at a larger scale the user has no idea how valid the classification technique is at that scale.

Calculating Errors

Calculating the total error by combining all these errors together can be difficult. It is hard to know if one error type dominates, or the total error is an equal combination of all types. Milford (1992) suggests that the total error factor is not the sum of the individual errors, but the square root of the sum of the squares of the individual errors i.e.

$$\text{accuracy} = \sqrt{(\text{error}1^2 + \text{error}2^2 \ldots\ldots + \text{error}n^2)}$$

For example, If you were digitising a 0.25 millimetre line on a 1:50 000 map with a high precision digitiser (0.025) the error factor can be calculated at:

$$\text{accuracy} = \sqrt{(1.25^2 + 12.5^2 + 5^2)} = 13.5 \text{ metres}$$

This suggests that the true position of the line is within 13.52 metres of the digitised line.

The calculations of errors should be flexible enough to allow for any other qualitative errors that are applicable. For example, a geologist may feel that his geological interpretation only has a accuracy of 25 metres. This figure can be incorporated into equation also.

$$\text{accuracy} = \sqrt{(1.25^2 + 12.5^2 + 5^2 + 25^5)} = 28.4 \text{ metres}$$

A calculation for each data set in the Cowaramup GIS project was carried out with (where possible) consultation with the suppliers of that data set. Where any of the error components are not known for a particular data set it was assumed that the unknown error component is the same as the greatest error factor of that type. Table 2 shows the error factor calculated for each data set used in the project. Each data set has description file in the database which describes it's source, the method and scale of capture and the calculated error factor.

DATA	ERROR
SPOT panchromatic image	Unknown
Contours	7.5 metres (DOLA 1993)
Bore hole data	50.1 metres
Slopes	7.5 metres
Geology	9.1 metres
Mining Tenure	18.2 metres
Cadastre	9.1 metres
Town Planning Zones	60.2 metres
Geological Sample Sites	9.1 metres
CALM Estates	100 metres
Leeuwin - Naturaliste Management Area	100 metres
Roads	7.5 metres (DOLA 1993)
Control Points	>1 metre
EPA Areas	18.2 metres
LGA boundaries	25.5 metres
Faults	375.0 metres
Coastline	25.5 metres

Table 2. Error factors for each data set

METHOD OF DISPLAYING ERRORS

Once we have calculated the error factor for each data set, how can we make use of this value? The system developed for the Cowaramup GIS project has two methods of displaying the error factor, by a textual report and by graphic error zones.

Error factor report

The first method of displaying error factors is by use of a error factor report. A report showing the scale of the display and the error factor for each data set is displayed on the screen. This also shows the maximum and average errors in both ground and screen units (See Table 3.).

ERROR REPORT

SCALE 1: 8897

COVER	GROUND ERROR	SCREEN ERROR
Bore	50.1 m	5.63 mm
Tenements	18.2 m	2.05 mm
Geology	9.1 m	1.02 mm
MAXIMUM	50.1 m	5.63 mm
AVERAGE	35.8 m	2.90 mm

Table 3. *Error Factor Report.*

This information can be useful to the user when constantly zooming in and out. Supposing you are interested in analysing geological information for a particular region, the usual way is to display (and possibly reselect from the database) the various data sets you require and zoom in to get the best possible view. Once you are happy with the extent of your view, you can request an error report which will tell you the scale of the display as well as the error factor in millimetres so you can gauge if accurate analysis can be made of the region. If the error factor is too great you may have to zoom out to a more appropriate scale to carry out your analysis.

Graphic error zones

The second method used to display errors in the Cowaramup GIS project is to graphically portray the zones of uncertainty for each line on the screen (See Fig 3). This method makes use of the scale of the display and the symbology function of the computer to display the width of each line according to it's error factor. This gives the user and instant appraisal of how accurate the data is at that particular scale. It also provides a confidence buffer when overlaying two or more data sets. Two data sets, for example, may appear to not to overlap, but when the error zones are displayed they do overlap slightly. It's highly likely that the two data sets do not overlap but this system will highlight the fact they may overlap and this information may be important for some purposes.

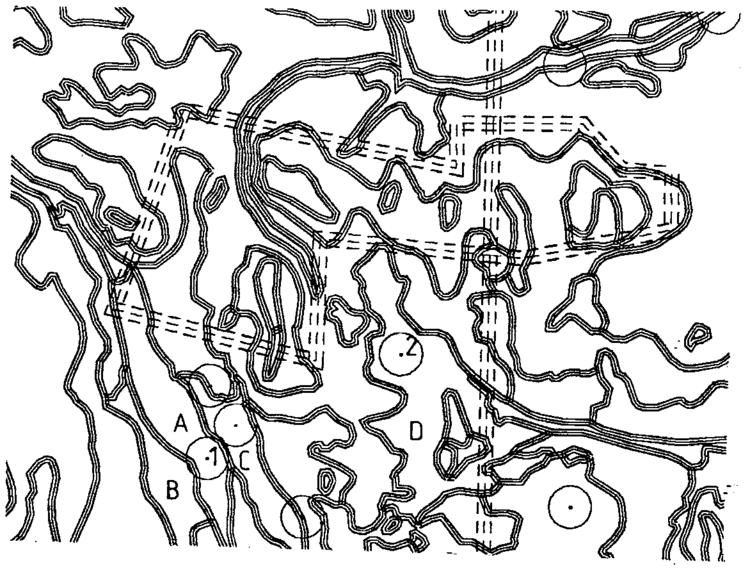


Fig 3. Three data sets, tenements (dashed line), geology (solid line) and bores (points) with their error zones . We can deduce from these that bore 2 is definitely geology type D, while bore 1 is possibly in geology type A, B or C.

This method displays the error zones as straight lines parallel to the digitised line. This is not the most mathematically correct method of displaying error zones, but the intention of this system is only to highlight that there is a error factor in the data set and to graphically display an approximation of it. Dutton (1992) discusses an model for positional error whereby the error zones are a sausage shaped locus around the line. This a more mathematically correct model for displaying errors and probability contours can be calculated so that points that fall close to the edge of a polygon can be said to have certain probability of being in the polygon. This would have the advantage of controlling the confidence level of the data to be analysed but would required complex algorithms to be formulated and would increase storage requirements as much as fifty percent for a two dimensional data set.

MINIMISING ERRORS

As mentioned before digitiser operators can introduce errors, especially if they are inexperienced. Digitising is one of the most monotonous and tedious jobs associated with GIS and is usually carried out by the more junior (which usually is the most inexperienced) people in the group. If accurate digitising (and therefore accurate decisions) are required by your organisation then experienced digitiser operators are essential components of your GIS team. Chrisman (1990) suggests that a crucial part of data capture is to select the right personnel and that operators should also be responsible for editing their own work therefore creating an incentive for accurate digitising.

Having a policy to digitise only source material with a scale larger than the scale of your intended output will ensure that errors will be minimised. The GIS office in the Department of Minerals and Energy have such a policy, but it is virtually impossible to complete a project with all of the source material available at an appropriate scale. If the appropriate source material is not available then region may need to be resurveyed at the required scale. This is usually too expensive, but may be cost effective if the project depends on that level of accuracy.

Including Meta data when supplying data to other organisations is a important consideration. Information such as date, scale, method of capture as well as type and the co-ordinate system of the source material should be offered to the receiving party if the data is to be used in a meaningful way. This information can be in written document form or it can be coded into the data itself. The Department of Minerals and Energy maintains a description file for each data set which is automatically picked up when exporting data to outside organisations.

ANALYSIS WITH ERRORS

The advantage of GIS over conventional forms of mapping is the ability to overlay and intersect data sets and to analyse the relationships between them. If those data sets have errors in them it follows that the results will have similar errors.

The GIS system when overlaying and intersecting data sets, will treat the boundaries as absolute. This may produce peculiar results such as sliver polygons. Features that are supposed to abut will overlap or have gaps which will in turn create very small areas of ambiguity. This may be a positive result in so much as it will highlight areas for further investigation. On the other hand, when doing point in polygon analysis some points outside the polygon but inside the error zone will be excluded. Another problem area is what error factor will a derived data theme have, taking into account the error factors of the themes used to create it. The policy of the Department of Minerals and Energy is to give the derived data theme the same error factor as the creating theme with the largest factor.

How then, do we allow for these errors? Using the system described above for the Cowaramup GIS project we can produce an error report of the particular data sets we are interested in. If the error factor is acceptable then we can use the machine to go ahead and perform the analysis. If the error factor is to great then an on-screen survey (using the graphic error zones technique) is required to determine if there are any features that are within any other feature's error zone. These problem features need to be recorded for further investigation. It may be the best policy to try and include all features that fall within error zones. The points and polygons can be buffered to their error factor width and the analysis can then be carried out. This would ensure that points that are possibly in the polygon are treated as totally in the polygon.

The user should be able to carry out analyses on the data with some idea of the error factors involved. The error report and error zone techniques provide the tools to achieve this. Although the precise error of the features in the data set is difficult to ascertain, this system gives the user some insight to the accuracy of the data.

CONCLUSION

There is no cartographic system that is free from error whether it is a paper map or a state of the art Geographical Information System. The Earth is far to complex, and to portray it on some medium accurately can be a costly or time consuming (or both) exercise. Somewhere a happy medium must be struck between the accuracy and the cost of that data. Very few organisations have the resources to go out and capture all the relevant data required for a GIS project. Most of the time data is available in hard copy form or held digitally in another organisation. Using this data may introduce errors that are not directly controllable, although they may be able to be calculated. This paper described some of these type of errors and suggested ways of allowing for them when analysing geographic data. The time will come where software vendors will incorporate error reporting facilities into there systems, but it will still be up to the user to recognise that these errors exist and to appreciate them.

REFERENCES

Burrough, P. A.
1986 Principles of Geographical Information Systems for Land Resource Assessment. Oxford: Clarendon.

Chrisman, N. R.
1990 Efficient digitising through the combination of appropriate hardware and software for error detection and editing. in Peuquet D J and Marble D F (ed.) Introductory Readings in Geographical Information Systems. London: Taylor and Francis: pp 165-174

Chrisman, N. R.
1991 The Error Component in Spatial Data. in Maguire D J, Goodchild M F and Rhind D W (ed.) Geographical Information Systems Principles and Applications Vol 1. New York: Longman Scientific and Technical: pp 233-245

DOLA (Department of Land Administration, Western Australia)
1993 (Draft) Data Users Guide Vector Topographic Data for GIS (all resolutions)

Dutton, G.
1992 Handling Positional Uncertainty in Spatial Databases. Proceedings 5th International Symposium on Spatial Data Handling Vol 2.

Milford, K. S.
1993 "Co-ordinate Systems, Map Projections and GIS." in ControlAltDelete First Quarter Vol 3 No 1: pp 10-15

FLOOD MAPPING USING G.P.S

Malcolm Ryan
Chief Planner
Byron Shire Council
P.O. BOX 159,
Byron Bay, N.S.W. 2481

Darryl Clark
Director
Geo-Spectrum (Australia) Pty. Limted
Suite 45, 2 OÇonnell Street
Parramatta, N.S.W. 2150

John Devine
GIS Support Analyst
Datamation Software Systems Pty. Ltd.
Suite 102, 88 Christie Street
St.Leonards, N.S.W. 2065

ABSTRACT

One of the needs that has arisen in Council's area was to accurately determine land affected by flooding and the affect on development this would have. Council does not have available any accurate contour information on which to base this flood mapping, although, detailed flood studies have been carried out to determine the height of the expected flood waters for the various catchments.

To overcome this problem new low level colour aerial photograghy was used as a vehicle to photogrammetrically generate one (1) metre contours at a scale of 1:1,000. Global Positioning System (G.P.S.) techniques were used to establish the photo control at pre-targeted locations. This was converted to digital data and translated to Council's mapping system to produce accurate contour information to overlay Council's cadastral base.

Council is now able to prepare maps showing properties affected directly or indirectly by flooding. Property records are updated with appropriate notations to ensure that any certificates produced, clearly show the extent to which the property is or is not affected by flooding.

This, we believe, is one of the first examples of the use of aerial photography and traditional photogrammetric techniques linked to G.P.S. to increase their accuracy and supply data to a land information system to help produce accurate information concerning flooding in the urban area of Byron Shire.

INTRODUCTION

This paper deals with the use of Global Positioning technology to prepare maps from Council's land information system. The purpose of the exercise was to produce a map showing land affected by flooding for the purpose of Council's Zoning Certificates (Sec. 149 Environmental

Planning and Assessment Act, 1979). The difficulty that Council has currently, and in the past, is the existing orthophoto maps are not survey accurate and the existing flood information is given to Council via its Consultants in height only (i.e. A.H.D.). The maps supplied by the consultants are taken from 1:25,000 topographic maps and therefore inaccurate in terms of the exact boundary of flood affected land.

Facing this difficulty and the fact that the contour information available to Council by orthophoto maps is very old and incomplete for various parts of the Shire, we investigated the possibility of using a surveyor to produce survey accurate contour information based on new aerial photography. The local surveyor, Bob Canty, of Canty and Williamson, Surveyors was keen to test his ability to produce this accurate contour information in the most cost effective manner. He considered that by using global position system hardware he would be able to halve the cost of survey when compared to conventional traversing from a known point.

Council also contracted with Geo-Spectrum (Australia) Pty Limited of Sydney for photography and production of the contour map using information supplied by Bob Canty. This has resulted in Council receiving for three urban areas of the Shire, survey controlled one (1) metre interval contours which can be superimposed over the existing cadastral base.

HOW WAS THIS DONE

Geo-Spectrum in conjunction with Bob Canty carried out a digital topographic mapping programme over an irregularly shaped area within Byron Shire, known as Ocean Shores Estate, as part of a Flood Study conducted by Council.

The specification called for the generation of one (1) metre contours, supplemented by a reasonably intense network of spot elevations in the flatter areas, to ensure adequate definition of the terrain.

This mapping also showed major relevant planimetric features such as the basic road system and all drainage features, and was presented as stable base inked plans at a scale of 1:1,000, together with the feature-coded digital data.

Colour aerial photography at a nominal scale of 1:8000 taken with a Zeiss LMK aerial camera (with forward motion compensation) and a 152mm lens, was used as a vechicle for the mapping.

Targets were placed prior to photography at the great majority of the 24 control stations established, the balance being at existing features that were clearly identifiable on the imagery. These targets were painted onto either a black plastic background, to enhance their definition on the photography, or directly onto the roadway. This gave an average of two (2) control points per stereo-model. Additional height only points were established in some areas to enhance the height definition.

Bob Canty then, with the assistance of a global position system unit, (Leica 200) co-ordinated these points. The GPS unit was hired together with its operator, one unit being placed on a trig station in the Ocean Shores Estate, the trig station having a known height above sea level together with ISG co-ordinates at full survey accuracy. The second unit was transported round the site over to the co-ordinate point. The Trig station acted as a base station and recorded continuously so that the fluctuations in the satellite were measured continuously against the known reference point. The second unit was roamed over the site to each point marked by the tape or otherwise. Acquisition time was between 5 and 15 minutes at that time, however, Bob Canty is now advised acquisition time is around 5 minutes given to more accessible satellites. The X, Y co-ordinates were accurate to 1 cm however the Z co-ordinates were less accurate. This data was converted from a geoidal reference to a flat plane reference and its accuracy was upgraded by a transformation equation supplied by the University of N.S.W., reinforced by spirit level connections from known bench marks. Geo Spectrum then densified this control network using the Bundle method of aerial triangulation to facilitate the digital mapping phase.

THE FINAL STEPS

The final maps were produced using Geo Spectrum's Computer Aided Resource Information System (CARIS) operated on the mapping front end of a Micro Vax II machine. Feature coded digital output was supplied on diskettes in DXF format and converted onto Council's mapping system, the Geographic Data Management System (GDMS), using utilities supplied by Datamation Software Systems Pty. Ltd. The conversion program went through without fault, aided by the clear understanding of precisely how the author of the data had set their layers, (this was one of the requirements on Geo Spectrum). Total data received so far is approximately 7 megabytes.

USING THE DATA

The information received by Council has been placed in separate layers on GDMS as a series of line strings. Information received included the contour information, form lines (estimated contours under the tree cover), the kerb and gutter of roads, watercourses, tracks, water bodies, coastline and the spot heights. Hard copies of the maps were also supplied on stable base from Geo Spectrum (examples of these are available for viewing). These contours and form lines will need to be joined and polygonised where possible, then tagged with the relevent height information. Datamation is currently investigating the use of a third party proprietory package to triangulate this information so that slope and aspect maps can be generated using this contour base together with the spot heights.

The next step is to analyse the flood height information supplied by our flood study and produce polygons of flood prone land. This consists of two polygons, one for flood prone lands and one for minimum floor levels.

It was interesting to note that when the contour information was overlaid on Council's existing cadastra it revealed just how accurate the existing cadastra was. This can also be seen from drawings supplied showing that the road reserves and the kerb and gutter lines from the two different sources coincide remarkably well. This gives me great heart to realise that although the cadastra is not technically survey accurate, it is of a fairly high accuracy and the subsequent flood maps which we have now produced will be reasonably accurate.

INTERGRATING CORPORATE DATA

Byron Shire Council currently operates a land information system based on a relational data base (MUNICS) linked to the GDMS mapping system, to cover all aspects of Council's functions.

Council contracted Datamation to prepare a series of programs to update Council's property system from the mapping system.

- Firstly, a program was needed to create notations against properties. The basis of this program centers around the establishment of a polygon in GDMS such as the flood prone land as outlined above. All properties that are wholly or partly within this area, are identified and the associated notation information, is recorded against the relevent property records in MUNICS. Council will be utilizing this program to place notations against properties affected by flood prone lands, bush fire hazard, slip prone lands and any other constraint that is mapping based. This is required for producing Zoning Certificates known in New South Wales as Section 149 Certificates.

- Secondly a program to produce buffer zones generally around a point source such as 200 metres from cattle dip sites, 1 km from piggeries or 1 km from quarries. Here again the mapping system will be used to locate the specific feature and generate the buffer zone and notate all properties within that buffer.

- The final program is to update land use from the mapping system. This program will enable a land parcel to be identified and a selection made from a table of land uses. The land uses are updated on to the MUNICS property system.

GENERAL FEATURES OF NOTE

AERIAL PHOTOGRAPHY

In this instance, despite a number of attempts to fly the photography, an unusually prolonged and frustrating period of poor weather delayed the acquisition of the imagery by over two (2) months.

TARGETS

The security of some of the targets in the more vulnerable areas, particularly along the beachfront, was a problem and they had to be checked/repaired/replaced for each photographic attempt.

GPS CONTROL

As the project area was mainly flat with considerable vegetation, and hence usually no line of sight between control stations, GPS gave an efficient, quick, and cost effective method of co-ordinating the control.

MAPPING

In some sections of the project area, the vegetation was too dense to accurately show one metre contours, and in these cases "formlines" were indicated, together with spot heights at locations where the ground was visible.

Height checks by ground survey were undertaken in clear terrain in selected areas in the project area and, apart from isolated instances, agreement of better than 10cms was achieved.

The photogrammetric approach allowed generation of the necessary information, to suitable levels of accuracy, for the flood mapping in an area where traditional ground survey techniques were both impractical and more expensive due to access and line of sight restrictions, and overall project size.

GENERAL

The method used on this project can, and has been in other areas, extended to provide 3-dimensional data to a very acceptable level of accuracy on all relevant natural and cultural detail that is discernible on the aerial photography, to give a more comprehensive data set.

Such information would include buildings, roads and tracks, utilities, vegetation, fences, pools, recreation facilities (golf courses, playing fields, etc.).

SUMMARY

As can be seen from the above, the direction Byron Shire has taken at the present time is towards the automatic production of Section 149 Certificates and their continual updating via our mapping system and to use the mapping system to carry out planning investigation studies as part of their ongoing forward planning programs. The by-product of the contour mapping is obviously very new and accurate aerial photography. Council proposes to continue with a flood mapping program with the extension into the Byron Bay township, Mullumbimby and Suffolk Park areas in this year's and next year's budgets. The surveyor assures us that the processing of these areas will be straight foreward now the methodology has been established. Council looks forward to developing the use of these techniques to other applications in the future.

LEVERAGING THE BENEFITS OF GIS AS A CORPORATE INFORMATION SYSTEM

Terry Ryan
Manager, Planning and Strategic Services
Information and Communication Services Division
Water Board Sydney, Illawarra and Blue Mountains

ABSTRACT

This paper gives a management overview of the business issues that affects delivery of the benefits of Geographical Information Systems (GIS). The paper suggests that the benefits of this technology are optimised when there is a clear business focus involving all dimensions of the business. The paper identifies business drivers and underlying customer demands that will either make or break a successful implementation of GIS. The importance of GIS as part of an integrated business information strategy is emphasised in relation to its role as a stand alone network, organisational or an enterprise technology solution.

It is shown that GIS is a strategic technology that can provide a competitive advantage in the delivery of integrated information products satisfying specific customer requirements and providing an opportunity for commercial returns.

Leveraging all the benefits of GIS requires a holistic approach involving people, organisational, business, quality and technology considerations.

INTRODUCTION

GIS applications within the Board can be described as falling within the following business and associated system development category as shown by **FIGURE 1.**

FIGURE 1: GIS APPLICATIONS

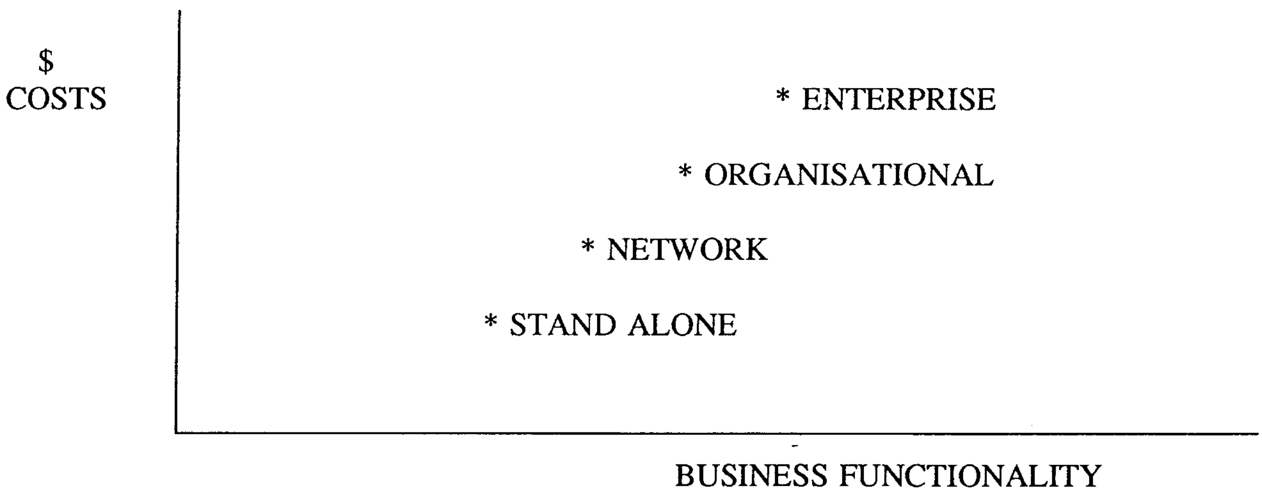


FIGURE 1 is a subjective assessment of GIS technology in relation to relative costs and business functionality associated with the *reach and range* of GIS applications. It is suggested that GIS *reach and range* may be described as stand alone, network, organisational or enterprise technology. These categories may be further expanded and related to organisational considerations:

- project specific,(stand alone)
- business operations,(network)
- programme delivery,(organisational)
- corporate support,(enterprise)

Experience has shown that costs are proportional to business functionality and the type of GIS application. It can be seen that successful GIS applications depend on an approach involving the above categories as well as the following inter-related factors which are components of the *big picture* (Ryan and Masters, 1991):

- organisational
- business
- people
- quality
- technology

Experience continues to show that the benefits are proportional to the degree of implementing the *big picture* and the synergistic product and service opportunities that comes from the integration of information systems and resultant value-added outcomes.

Too often, the time and attention needed to manage individual projects obscure the need to evaluate the organisation's total need. The inability to see the *big picture* often works against optimising GIS benefits.

It has been observed that where GIS applications have been driven primarily by technology, the benefits provide limited business returns. Typically, GIS technology is selected for short-horizon projects with specific outputs and does not address the longer term corporate concerns. Often the technology usually does not survive beyond the life of the project and in some cases the application has been abandoned.

In cases where benefits have been quantifiable, there has been a specific business need, major re-engineering of work processes, organisational changes, continuous productivity savings and most importantly, improved people skills. People skills ultimately provide the operational *know-how*. Making it happen depends on organisational and people commitment and are perhaps the most overlooked factors in optimising the benefits of a GIS implementation.

Overall, GIS benefits have ranged from tangible productivity gains to nebulous intangible benefits or in extreme cases, costly non-returns.

ORGANISATION

In common with many government trading enterprises and utilities around the world, the Board is undergoing major organisational restructuring. The restructuring is aimed at transforming a statutory authority into a commercially productive and profitable organisation, providing a high level of service to its customers.

As part of this process traditional hierarchical processes have been replaced with operating and business areas that have well defined responsibilities and are aligned by corporate goals and key result areas. The Board has three operating areas involving Corporate, Core and Australian Water Technologies. Each operating area has business areas or units.

Information and Communications Services has been set up as a business area providing cost competitive services to its customers throughout the Board.

The transition is being driven by a *back to basics* approach that has focussed on simplifying core business delivery, customer demands and value for money. From an information technology (IT) perspective the impact of this may be interpreted as:

- simplifying a complex organisation
- distribution systems more reliant on communications (eg just-in-time, remote control and monitoring)
- increased use of spatial information (eg asset valuation, modelling, system performance)
- greater emphasis on process co-ordination as distinct from process refinement

Clearly a new business structure will require a new IT architecture.

In the context of the organisational direction, GIS technology must be able to operate within an open information system environment with an ability to support specific projects, business operations, programme delivery and corporate support. It is my understanding that spatial information technology within the Board will involve a number of GIS platforms, capable of handling generic data and integrated with strategic aspatial systems. In this context, GIS has a major role in providing total business solutions.

Total business solutions can be described as business processes that are cross-functional, product focussed and customer packaged. The customer prefers to deal with only one point of contact and is generally not concerned with technical or big picture issues. Timeliness, cost and quality are the major customer issues.

BUSINESS

IFIS (Integrated Facilities Information System) is a corporate (enterprise) information system involved primarily in data capture, maintenance, upgrade and analysis of property and services information. The data capture operation began in 1985 involving the digitising of 30,000 hardcopy maps and completed early 1993 with the following outcomes:

- 1,200,000 land parcels
- 20,000 kilometres of water mains and fittings
- 21,000 kilometres of sewer mains and structures
- 400 kilometres of stormwater channels and drains

The main customers for IFIS products are the Board's Regional businesses with a growing demand from other operational and corporate areas which include: Water, Waste Water, Planning and Finance. The business drivers are:

- customer enquires on location of services, property identification and house services diagrams
- base information for planning of service extensions and protection of existing assets
- location and networking of mains, valves and manholes
- capture of asset attribute data (date constructed, asset type, length of asset) for asset revaluation and determination of asset depreciation
- data on wastewater assets in relation to planning, modelling and analysis of operating systems
- integration with the monitoring and remote control of the Water network

The major benefits have come from productivity savings involving:

- reorganising the structure and work classifications with emphasis on GIS technology and moving away from the traditional survey drafting role
- introduction of project team and multi-skilling operations, flattening the organisation structure and pushing decision making down
- downsizing the workforce
- improved accessibility to data through Board wide inquiry system
- reduced service disruption calls
- potential environmental impact of new work
- improved access to base data for capital works
- hard copy plan production
- improved data maintenance processes
- maintenance operational costs reduced
- revenue maximisation
- purification of customer database information

IFIS benefits to the Board have been significant, however their visibility have often been obscured by intangible returns, reluctance to change traditional work practices or cultural attitudes towards information technology.

The Board has also invested in other GIS applications which include:

- Clean Waterways Programme
- Environmental Management
- Urban Planning
- System Planning
- Catchment Services

The following **TABLE** profiles these applications:

TABLE: GIS BUSINESS AND TECHNICAL PROFILES

BUSINESS		TECHNICAL	
AREA	APPLICATION	ISSUES	GIS PRODUCTS USED
CLEAN WATERWAYS PROGRAMME	Strategic planning and modelling of board's wastewater and stormwater systems	Ongoing life of project; Flow of data between systems	INTERGRAPH (INTERGRAPH UNIX workstations and PC's)
ENVIRONMENT MANAGEMENT UNIT (EMU)	Various environmental and water quality projects	Flow of data from other systems and organisations	ERMS & SPANS (PC's)
URBAN DEVELOPMENT	Transfer of corporate data from IFIS and proposed developer data to IFIS; some environmental projects	Interfaces with CADD systems; migration to new IFIS platform	GEODIS & SPANS (IBM UNIX Workstations)
SYSTEM PLANNING	Strategic assessment of asset quality	New IFIS Platform	MAPINFO (PC's)
CATCHMENT SERVICES	Catchment Management	Assessment of new technology for all of Bulkwater	ERMS (PC's)

Further benefits have been gained by other organisations who have licensed IFIS data in setting up their own GIS operations. The savings to these organisations were cost justified and significant.

BUSINESS DRIVERS

The strategic business drivers for GIS as a component of the broader information technology requirements for the Board are:

- corporate business objectives
- information systems must continue to change to facilitate commercial transactions
- a balance between local and corporate management of information
- need to accelerate change in cultural attitudes to gain optimum benefits from information technology
- help customers get their business requirements right
- total business solutions
- greater involvement of customers in the design and modification of systems

The key business drivers for GIS activities producing commercial products include:

- understanding customer demands
- quality products and services
- ensure that technology is aligned with customer's business objectives
- commercially competitive
- supply of products at agreed price, timeliness and quality
- meet growing demand by linking changing business requirements to emerging technologies

QUALITY

The notion of quality as it applies to GIS ranges from technical measures of spatial accuracy, data integrity, topology, source, input processes to subjective customer satisfaction measures.

In delivering GIS products, the service provider must also be aware that the customer's satisfaction includes technical measures as well as the following business considerations:

- customers believe that GIS have matched their requirements
- customers are treated right
- quality of communication in business terms and appropriate technical language
- value for service
- 100% technical performance

The increased use of geographical information has also emphasised the quality of the information. This has led to circumstances where more time is spent on quality assurance processes than ever before.

In this light, it is noticeable that quality management procedures already consume a large proportion of the data input process. Corporate moves to "Total Quality Management" and the expected requirements for improved quality documentation, imply that much more time will be spent on quality maintenance processes in future.

Quality from a GIS perspective, is not just about positional accuracy but embodies a host of perceptions and performance criteria. Measures of quality can therefore range from being very subjective to being extremely rigorous. Documenting quality is therefore a difficult process and covers many aspects apart from assessing positional accuracy and is a subject requiring greater attention (Ryan and Masters, 1991).

Quality is also increasingly perceived as being related to customer expectations of the system. The customer's perception of Quality therefore can be related to the performance of the system rather than the data itself. For example the perception of quality is affected by how long it takes to obtain the required information, or even whether the information can be obtained. However, data quality is also affected by data accuracy, completeness and conformance to specified requirements.

Quality Assurance accreditation is increasingly becoming a business imperative. GIS suppliers are faced with the decision to either establish recognised quality management systems or risk losing market share to accredited competitors. Customers have indicated a preference towards *quality* organisations and businesses have seen it as part of their competitive advantage.

As a supplier of GIS products, steps have been taken towards accreditation under the Australian Standard 3902, which has involved:

* documentation of work manuals
* policy
* business functions
* customer requirements
* management systems
* training
* quality teams

The outcome from quality management and steps towards accreditation has provided significant business benefits in relation to customer satisfaction, work processes and high quality products. The main benefit of quality management can be seen as providing a systematic way of guaranteeing that information, people, technology and business activities happen the way they are planned.

PARADIGM LOST

GIS must have the functionality to best fit particular business needs, however the *real GIS* benefits are not just technology, they come from sweeping changes in management and organisational structure that redefine how the work gets done.

Traditional work assumptions and procedures must be challenged and changed to meet a new work paradigm that redefines work processes that increases productivity, competitiveness and strategic value. The changes within the Board involving GIS operations have largely been driven by internal customer demands for greater value for money and improved operational performance.

GIS benefits do not come from whiz bang technology alone, they come because it supports break through ideas in business processes.

The real benefits come from focussing on changing the business through organisational structures, re-skilling people, smarter work processes and aligning GIS technology with specific business needs.

REFERENCES

Ryan, T and Masters, E, 1991. Spatial Accuracy as a Quality Indicator. Paper, Conference on Land Information Management, University of New South Wales.

GartnerGroup, 1993. Industry Service Research Notes, IT and Business Strategy: A Research Summary.

ACCOUNTING FOR THE COSTS

OF AGRICULTURAL PRODUCTION

R.E. Smyth, P.A. Walker,
CSIRO Division of Wildlife and Ecology
PO Box 84, Lyneham, ACT 2602

ABSTRACT

The SRIAS (State Resource Information and Accounting System) project was started in 1992 to examine broad-scale resource, environmental impact and economic policy questions in relation to New South Wales. It is being undertaken by CSIRO with assistance being provided from a number of NSW and Federal government departments and authorities.

As a first step, the project is attempting to evaluate, on an areal basis, the value of agricultural production and its associated costs.

The revenue side is accomplished by allocating ABS statistical production data areally using NOAA satellite information, pattern analysis and linear programming techniques. The expenditure side allocates transport costs, government subsidies and programs, and resource depletion costs also on an areal basis using a variety of techniques.

This paper examines an aspect of the expenditure side, one of the resource depletion costs - soil erosion, and attempts to come to a state-wide estimate. A GIS and various other spatial modelling tools are used. The techniques and possible improvements that may be made are also discussed.

INTRODUCTION

The idea of resource accounting is not new. Discussing resource management Heady (1966) observed that those scientists who construct 'soil nutrient balance sheets' fail to recognise the concept of inter-regional transfer. A small amount of resource depletion in one area to facilitate substantial improvement in another may result in a net gain to society.

Concerned about resource conservation others (Garbarino, 1992) prefer monetarised and non-monetarised accounting systems. These views represent two aspects of similar but related issues. Both seek means to improve social welfare.

Neither resource management nor conservation, however, can be examined in isolation of the other. What is needed is an approach encompassing both viewpoints; an approach which allows detailed analyses within a framework of a general overall picture. A picture that provides the context for a more detailed analysis and one that integrates social, environmental and economic information in a framework that facilitates the use of different methodologies.

SRIAS - a Statewide Resource Information and Accounting System (SRIAS) for the state of New South Wales - is our response to this need. Using the power of geographic information systems and focused at the strategic level, it is designed to allow consideration of a wide range of policy alternatives.

SRIAS - PROVIDING AN OVERVIEW OF RESOURCE USE

SRIAS - a State-wide Resource Information and Accounting System is our response to these issues. It is an attempt to provide a framework, which may be refined and improved over time, within which the issues of biodiversity and resource management may be examined.

SRIAS is focused at a broad level, both spatially and temporally. It is designed to allow us to consider a wide range of land-use alternatives and the impact of a range of resource use policies. It is designed to have economic and resource data that is spatially and temporally organised and internally consistent.

State-wide data in SRIAS cover diverse topics including land tenure, agricultural production, vegetation, geology, soils, population, degradation etc are integrated with annual data on agricultural production in both economic and physical terms.

SRIAS is a collaborative project between the CSIRO Division of Wildlife and Ecology, the NSW Land Information Centre (LIC), Bathurst - a part of the Department of Conservation and Land Management. Collaborative agreements have also been signed with the Australian Bureau of Statistics, other CSIRO and NSW government agencies, including the Department of Agriculture. Additional research agreements with AUSLIG and the Murray Darling Basin Commission are also in place.

In the following sections, we describe how SRIAS is being used to develop a 'resource balance sheet' for agricultural land use. As a first attempt at creating this accounting framework, a significant difficulty arose due to the lack of a realistic land-use map for the state. Such a map is essential to establishing the spatial extent of agricultural production and habitat changes. Obtaining a realistic land-use map, however, is not as simple as it might seem.

PRODUCING A NSW LAND-USE MAP

Agricultural land use accounts for 87% of the total land use of New South Wales. By contrast, the area currently in forestry and forest reserve systems is 4.9% and in nature conservation and recreation reserves is 4.3%. While SRIAS can be used to will be examining other land uses (eg. forestry see Walker 1992), our main focus is on agricultural land use. The first priority of the project has therefore been to attempt to produce a year-by-year picture of NSW land use.

Establishing a realistic land-use map for a state requires information on the location of roads, rivers, urban areas, forest and nature reserves and national parks. Beyond this, it is necessary to determine the type of agriculture production in remaining areas.

The starting point was a topographic base, including a river and stream network, provided by AUSLIG and the NSW LIC. These data were supplemented with data on urban areas (from Australian Bureau of Statistics) and forested areas and reserves (from the Land Information Centre NSW) and public lands (from AUSLIG). The remaining area was assumed to be under agriculture of some sort.

Determining the location of agricultural production

To determine the location and type of agriculture, analysis of AVHRR satellite data was used in conjunction with Australian Bureau of Statistics (ABS) production statistics. It was necessary to ensure that the land-use map was consistent with official production statistics and other policy models. Double cropping had to be removed and the area of agricultural activity made consistent with the area that is used for agricultural production.

Achieving this is an exercise in spatial disaggregation. More specifically, given an aggregated value for a region (in our case ABS agricultural production figures), the task is to re-distribute portions of that value spatially across the region such that the sum of the values associated with the locations within the region equals the total region value. In almost all cases, we know that the region is not homogeneous with respect to its contribution to the aggregated value.

In determining an appropriate methodology, we cannot assume an underlying smooth surface as required for contouring or Tobler's (1992) pycnophlactic interpolation. Flowerdew's (1989) method of spatial disaggregation was inappropriate since it returned a probability of a land use (not an absolute allocation) and was not constrained by a set of area totals.

Specifically, given an aggregated estimate of the area of each agricultural production (from ABS figures) within a local statistical area, out problem is to determine where did that production occur.

Our procedure involves the following steps :

1. Determine land use at key sites (a training set)
2. Determine the similarity of locations to key sites
3. Allocate locations to land-use types

Each step is discussed below.

1. Determine land use at key sites (a training set)

Our procedure for producing the land-use map involves processing satellite data (NDVI from 1kms AVHRR data) to establish growth profiles for the major categories of agricultural production. A growth profile consists of the sequence of NDVI values for 12 monthly images covering a single year. These profiles are, in most cases, distinct.

Cotton, for instance, has a quite different profile to grazing or cereal production. Problems arise, however, when profiles are similar. A well treed urban area can look very similar to a forest plantation, for instance. Similarly, orchards look very similar to irrigated areas, and canola is almost identical to a cereal crop.

To assist us with establishing the relationship between the satellite data and agricultural production, CSIRO and the NSW Department of Agriculture are developing a Register of Rural Land-use Sites (RRLUS). The register consists of key sites, each time-stamped for the type of production that has been determined. Details on the location (AMG and geographical co-ordinates), date, locational accuracy, land-use classification and confidence in the assigned of land-use class and the source material (eg. high resolution satellite data, ground survey, aerial photography) are stored for each key site.

These sites form the starting point (or seeds) for identifying agricultural production locations.

2. Determine the similarity of locations to key sites

A land-use map can be prepared by comparing the growth profiles of any location in New South Wales with those of the locations stored in the Register. Locations, in general, can be assigned the land use of the key site (in the register) which they are most similar to (comparing profiles).

The success of this process, clearly depends on the availability of information on the key sites reflecting the types of agricultural production in an area.

Key sites for each type of production were identified within each of Laut et al's (1980) Provisional Environmental Regions of Australia (PERA). This ensured, for instance, that if we were looking for horticultural production in an area (based on ABS figures), that we had to have a key site for horticulture in it's relevant PERA region for comparing growth profiles.

The key sites provide a starting point for a non-hierarchial clustering algorithm (Belbin 1990), which determines the similarity between the locations in the register and all other locations.

3. Allocate locations to land-use types

The identification of production areas, however, involves more than just measuring the similarity between each location and a set of production growth profiles.

To be compatible with government production models, the amount of land associated with each type of agricultural production needs to be constrained.

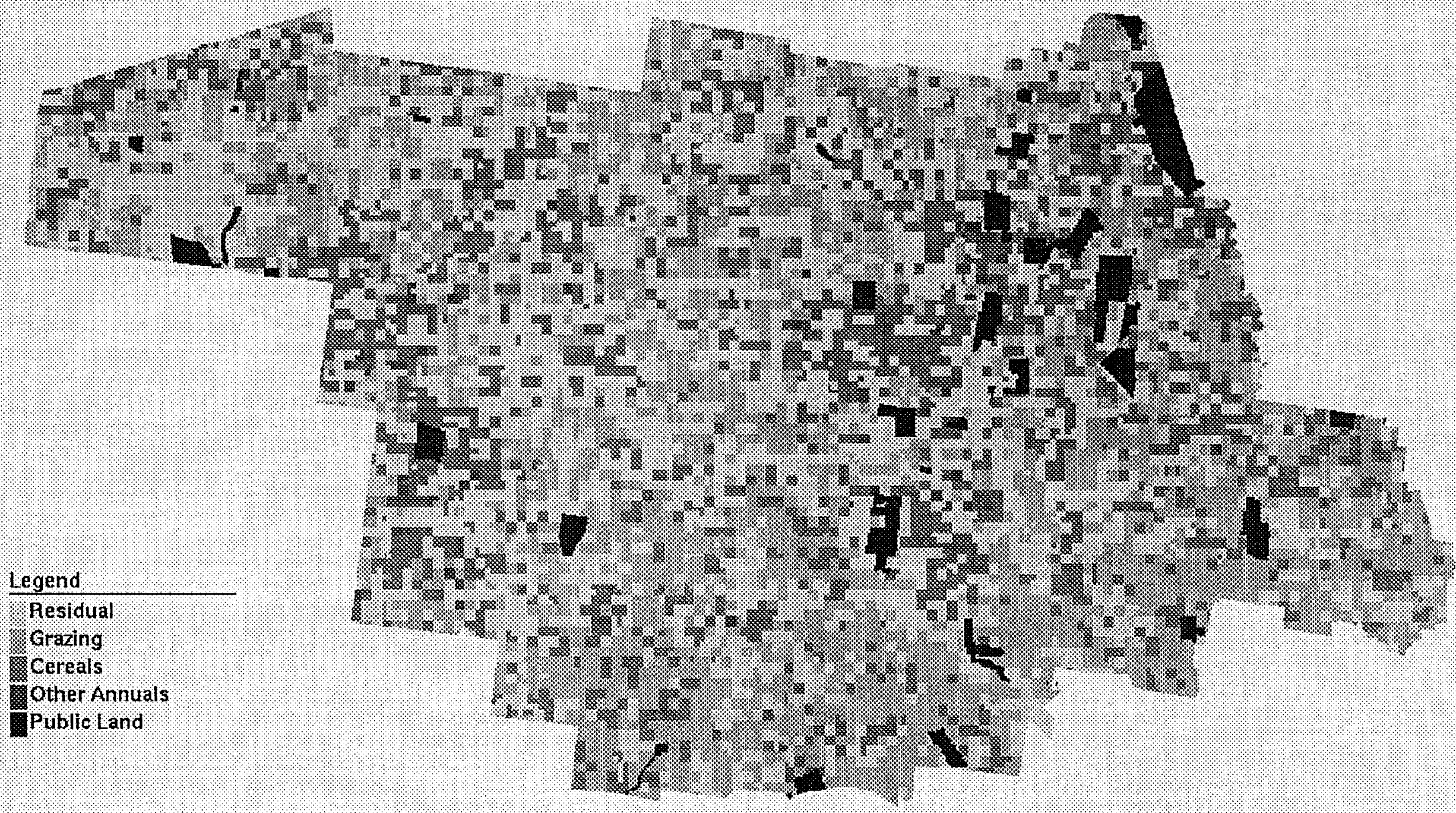
Our method developed involves determining the land-use allocation in which the similarity of the growth profiles of each location to each key site is maximized. This allocation is subject to the constraint that it can only allocate production to agricultural areas (ie. not public lands, state forests etc), and can only allocate as much production as the available statistics indicate occurred (ABS figures). Our methodology uses linear programming.

To date, we have succeeded in re-distributing ABS production statistics collected by administrative regions to a set of point locations within each region. The maps we have produced show a realization of a land-use map which is internally consistent with official production statistics and which conforms to our growth profiles. Land-use maps for 3 administrative areas have been produced and an example for Bland statistical local area is shown in Figure 1. They are in the process of being examined by the Department of Agriculture.

It should be stressed that these land-use maps need considerable ground truthing. However, given an examination of higher resolution satellite imagery revealing paddock level detail, at this stage the maps seem plausible.

Once we have completed that step, we plan to automate the process and extend it to all the Lachlan catchment, and ultimately the entire state of New South Wales.

Fig 1 – Modelled Land Use



DEGRADATION - DETERMINING ITS EXTENT.

Having prepared land-use maps for several years and across the entire state, the next step in the project is to be able to determine agricultural land at risk due to degradation and how this will change over time with differing land-use patterns, government programs and climatic events.

Recently, Walpole et al (1992) estimated that losses to the farm community from land degradation amount to $400m per annum for New South Wales. If these losses are to be accepted, then a serious look at the spatial implications of degradation across the state needs to be conducted.

There are however, a number of studies that have addressed this issue. The significant studies by NSW Soil Conservation Service are briefly outlined below.

Work undertaken by other researchers

In 1987-88 the NSW Soil Conservation Service published a Land Degradation Survey of NSW (Graham, 1987). Its aims were to assess the status of land degradation in NSW and identify the location, extent and severity. The forms of land degradation considered were:

- erosion (sheet, rill, gully and wind)
- mass movement
- soil structure decline
- salinity (irrigation,dryland seepage,scalding)
- induced acidity
- woody shrub infestation

The methodology did not incorporate measurement of any decline in productivity associated with degradation and the report states that there is scope to superimpose such assessments on the results at a later stage.

The methodology was a point in time survey based on sample points at 5 and 10 km intervals across the State. Assessments of degradation at each sample point were made using aerial photographs, specialist field knowledge and existing data. Sheet, rill and wind erosion were assessed on the basis of erosion hazard, viz., the relative level of erosion with current land use. This was computed using a modified form of the USLE (Wischmeier and Smith, 1978). The other forms of degradation were assessed on the basis of observed degraded status with no attempt to predict future occurrences.

This data set has been kindly made available to the SRIAS project and will be most useful for model validation.

In other work undertaken by the Soil Conservation Service, Edwards (1987) details runoff and soil loss studies undertaken in plot trials which first started in NSW in 1943. The major features of these studies were:

- relatively low soil losses by world standard
- high spatial variability of soil losses
- high temporal variability of soil losses
- the major effect of management practices

The low soil losses identified in these studies need to be understood in the context of the close to zero soil formation rate in this country. Essentially Edwards asserts that any soil loss is significant.

Edwards argued that the USLE has limited application to Australia because of doubts as to the suitability of some of its factors and the difficulty of estimating others for Australian conditions. It should only be used for supplying relative measures.

McBratney (1992) also highlighted the effects of uncertainty in the USLE parameters on the final estimates for annual soil loss. Given uncertainty of as much as 50% in Erodability, 25% in Erosivity and Cover and Management practices, and smaller errors in the other parameters, he calculated that the resulting estimated soil loss could have an error of + 66%, and that moderate correlations between the components of the USLE not only effect the uncertainty but the estimate itself.

Edwards recommends the use of models that predict soil loss on an event basis and that data from the plot experiments should be used in validating the models.

Problems and Progress

Given the foregoing it would appear optimistic to embark upon this next phase of the SRIAS project...the modelling of degradation. However given the urgency evinced by ANZECC in its "Draft National Strategy for the Conservation of Australia's Biological Diversity", the development of a forecasting capability in SRIAS covering degradation is essential.

It must be stressed that what we seek is an informed strategic overview or realisation of what is occurring, and not an accurate description of exactly what occurred in every hectare of the state.

The technique being followed initially to gain a broad over-view is to repeat the calculations done by Graham using the Modified USLE with the benefit of the new land-use maps described above. As in the USLE, the components of this are Erosivity (R), Erodability (K), Slope length (LS), a Land-use factor (C) and a Management factor

(P), with modified Erosivity and Crop and Management factors (as in Williams, 1983). It is expected that after some initial exploration with this technique, use of the PERFECT model (Littleboy et al 1989) will be examined.

This is not, however as easy as it may appear when estimates are required for extensive areas. The USLE was based on a plane surface of slope length 22.1 metres and uniform slope of 9%.

The outcome we seek is a revised sheet and rill erosion hazard map. In the following, we examine the problems of establishing estimates of the model components for the modified USLE (MUSLE) for the entire state. Note that at this stage, we have not accounted for crop rotations and assume a 22 year sequence of climatic conditions, not the actual sequence that occurred in 1989-90.

Erosivity has been determined by linear interpolation of the erosivity map prepared for NSW by Rosewell and Edwards (1988), in the same fashion as Graham.

Slope data has been provided by the Murray Darling Basin Commission for the Murray Darling Basin based on AUSLIG 1:100000 map data. These data appear adequate for state-wide modelling.

Slope-length, however proved a more difficult proposition. Wischmeier and Smith (1978) define slope length as the distance from the point of origin of overland flow to the point where either the slope gradient decreases enough that deposition begins, or the runoff water enters a well-defined channel that may be part of a drainage network or a constructed channel. Significantly, a change in land cover or a substantial change in gradient along a slope does not begin a new slope length. This means that both the top of catchments and stream networks must be delineated.

The stream network is based on data from NSW LIC. Catchment boundaries were delineated using software developed by the USGS EROS Data Center (Jenson, 1991) and the 2.5km DEM developed by Centre for Resource and Environmental Studies (CRES) at the Australian National University. This software had already been used successfully by Hutchinson and Dowling (1991) on this DEM. An example of the boundaries derived by this software may be seen in the map of the Lachlan Hydrological Basin in Figure 2.

It is relatively easy to take grid based DEM points into the USGS program and import the output into a GIS (in this case SPANS). The output from the program was a grid of points containing a catchment number. This was imported into SPANS as point data, and a voronoi map produced of the catchments.

The major problem with our current approach has been the resolution of the state-wide DEM. Quinn (1991) states that distributed modelling of hillslope flows requires a grid scale much smaller than the scale of the hillslope. Unfortunately the elevation data used for computing the catchments is not of the same resolution as our slope map.

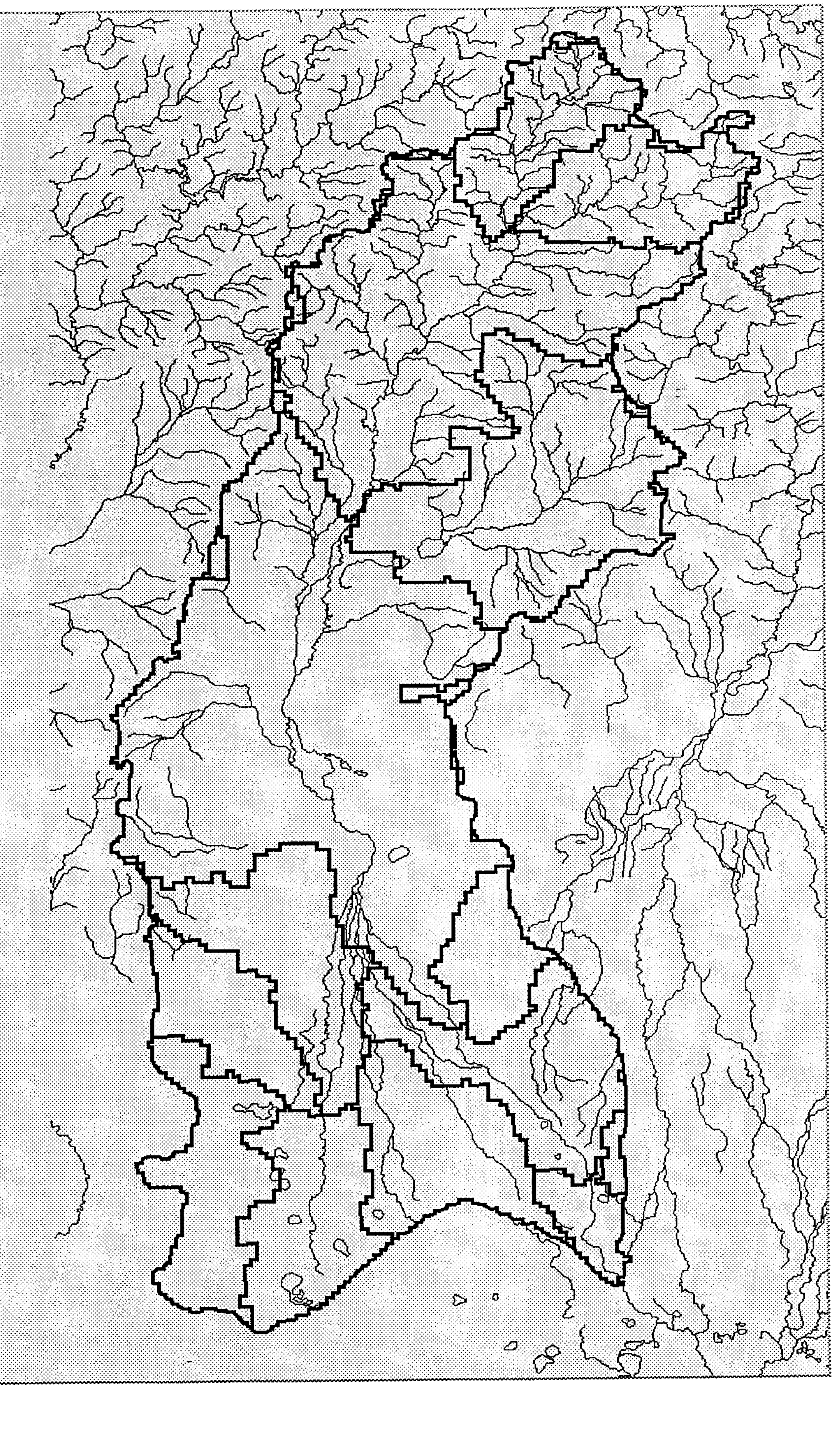


Fig 2 – Catchments in the Lachlan Basin

Initially we thought that the approach adopted by Zhou (1990) would be appropriate for calculating slope length, viz., the shortest distance from the point to the top of the catchment. Given the availability of software, it was decided that a modification of the USGS code would provide a satisfactory solution.

The work on how to define the slope length is proceeding and will be used in the second phase of the modelling when a process model is used. Initially, an approximation of slope length has been made and the Mutchler and Murphree modification for the exponent m has been used (Lal 1988).

The GIS (SPANS) was then used to append the MUSLE factors to each location in the DEM. The MUSLE Soil Loss was then calculated for each location. Note that is not a real soil loss, but a hazard estimate in Graham's terminology, due to the nature of the USLE. In fact it is best interpreted as an estimate of the average amount of soil erosion that would occur with that form of land use if was subjected to the climatic events that would be expected to occur over a 22 year period.

The resulting map (Figure 3) was produced as a Voronoi map and then classified using the same classes as Graham's original map, a section of which has been reproduced here from the Soil Conservation data (Figure 4).

Improving these estimates will require us to take account of rainfall events and a slope length index that gives us some estimate of deposition. As stated previously, the next step we are embarking upon is the process modelling of erosion, using rainfall event data from the Bureau of Meteorology and, it is hoped, better slope length data. We are also intending to try and take account of soil deposition as discussed by Moore, et al (1992).

The final phase of the project will attempt to map existing government programs and expenditures and relate this to the maps obtained in the previous phases of land use, productive capacity and ecosystems at risk.

Fig 3 – Modelled Erosion Hazard, BLAND
(CSIRO Map of Sheet & Rill Hazard)

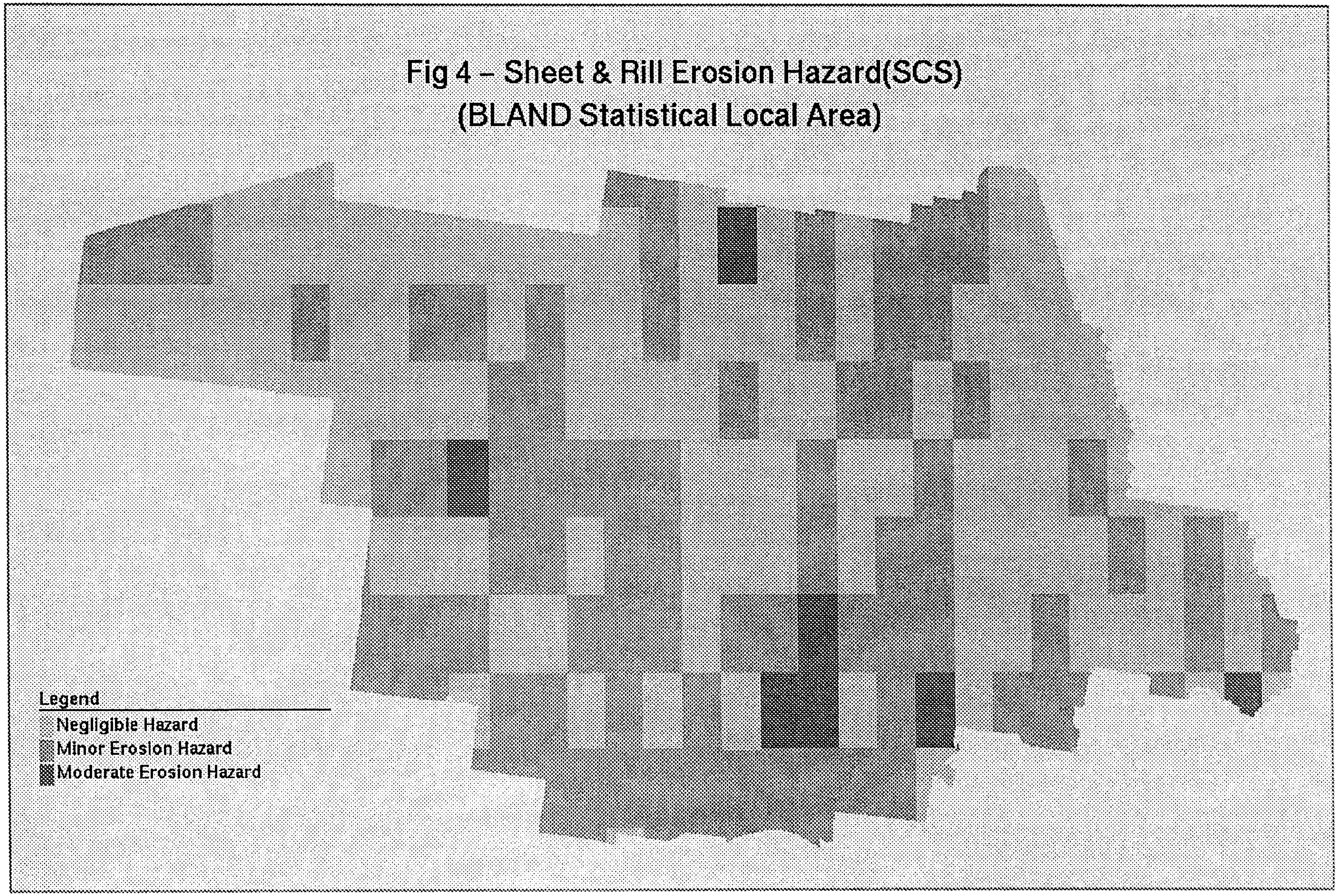


Fig 4 – Sheet & Rill Erosion Hazard(SCS)
(BLAND Statistical Local Area)
Legend
Negligible Hazard
Minor Erosion Hazard
Moderate Erosion Hazard

CONCLUDING COMMENTS

The establishment of a set of yearly land-use maps for New South Wales, each compatible with production statistics, is still some time off. The land-use component, however, is critical to any analysis of land degradation.

Even today, the question of where eroded soil goes is not clear, as Rose (1992) remarked in his summary of the erosion prediction workshop. With erosion we must ask ourselves how important it is to existing productive capacity and ecosystems. Very little is known about soil deposition. There are still significant problems in attempting to quantify many of the variables associated with degradation.

Despite this it is clear there is an urgent need to establish a framework in which the value of production, the costs of degradation and conservation objectives can be discussed. Such a system when developed will enable us to discuss the trade-offs in a spatial framework.

The steps we are undertaking are not easy, and not without risk. Our models and maps may be misinterpreted. The results may be criticised because of the data and models used, and because of the nature of the research and problems we are confronting. The main issue will be one of scale. We are concerned particularly that people will try to use the maps we are producing to derive conclusions about the way individual properties are managed. SRIAS is a system designed to provide the overview picture of land-use activities in New South Wales, and not for focusing on paddock or farm level details.

Apart from solving a wide array of empirical problems along the way one of the most difficult challenges that we face will be to find a way to present our data so that the errors embodied in it are appreciated. If we can cross this barrier, however, we consider that we will be in a position to make a significant contribution to resource management and prospects for ecologically sustainable development.

REFERENCES

Belbin, L. (1990), "PATN Pattern Analysis Package" CSIRO Division of Wildlife and Ecology

Edwards, K. (1987), "Runoff and Soil Loss Studies in New South Wales" [Sydney, Soil Conservation Service of NSW], A National Soil Conservation Program Project Conducted by Soil Conservation Service of NSW and Macquarie University.

Flowerdew, R., and Green, M. (1989), Statistical methods for inference between incompatible zonal systems. In Goodchild, M.F. and Gopal, S. (eds.) Accuracy of Spatial Databases. Taylor and Francis, London. pp 239-47.

Garbarino, J. (1992), "Toward a Sustainable Society: an economic, social and environmental agenda for our children's future" The Noble Press Inc

Graham, O.P. (1988), "Land degradation Survey of NSW 1987-1988: Methodology" Soil Conservation Service of NSW Technical report No:7

Heady, E.O. (1966), "Economics of Agricultural Production and Resource Use" Prentice-Hall

Hutchinson, M.F., Dowling, T.I. (1991), "A Continental Hydrological Assessment of a New grid-based Elevation Model of Australia" Hydrological processes, Vol 5, pp:45-58

Jenson, S.K. (1991), "Applications of Hydrological Information Automatically Extracted from Digital Elevation Models" Hydrological Processes, Vol 5, pp:31-44

Lal, R. (1988), "Soil Erosion by Wind and Water: Problems and Prospects" Soil Erosion Research Methods, Soil and Water Conservation Soc

Lane, L.J., Renard, K.G., Foster, G.R., Laflen, J.M., (1992), "Development and Application of Modern Soil Erosion Prediction Technology - The USDA Experience" Aust J Soil Res Vol 30 pp:893-912

Laut, P., Firth, D. and Paine, T.A. (1980), "Provisional environmental regions of Australia. A working document towards a framework for Australian environmental statistics.". CSIRO Melbourne.

Littleboy, M., Silburn, D.M., Freebain, D.M., Woodruff, D.R. and Hammer, G.L. (1989), "PERFECT: A computer simulation model of productivity erosion runoff functions to evaluate conservation techniques." Queensland Department of Primary Industries, Brisbane.

McBratney, A.B. (1992), "On Variation, Uncertainty and Informatics in Environmental Soil Management" Aust J Soil Research , Vol 30 pp:913-935

Moore, I.D., Wilson, J.P. (1992), "Length-slope factors for the Revised Universal Soil Loss Equation: Simplified method of estimation" J Soil and Water Conservation Vol 47, pp:423-428

Quinn, P., Beven, K., Chevallier, P., Planchon, O. (1991), "The Prediction of Hillslope Flow Paths for Distributed Hydrological Modelling Using Digital Terrain Models" Hydrological Processes, Vol 5, pp:59-79

Rose, C., (1992), "Erosion/Productivity and Erosion Prediction Workshop" Proc of the 5th Aust Soil Conservation Conference Eds Hamilton, G.J., Howes, K.M., Attwater, R.

Rosewell, C.J., and Edwards, K. (1988), "SOILOSS - A program to assist in the selection of management practices to reduce erosion.". Tech. Handbook no. 11. Soil Conservation Service of NSW, Sydney.

Tobler, W. (1979), "Smooth pycnophalatic interpolation for geographical regions." Journal of the American Statistical Association 74:519-530 .

Walker, P.A (1992), "Plantation Land Availability Study." Resource Assessment Commission, Forest and Timber Inquiry Consultancy Series. AGPS Canberra

Walpole, S.C., Sinden, J.A. and Yapp, T.P. (1992), "The opportunity costs of land degradation on agriculuture in New South Wales" Progress Report to the Land and Water Resources Research and Development Corporation, University of New England, Armidale

Williams, J.R. (1983), "The Physical Components of the EPIC model" pp:272-295 Soil Erosion and Conservation, Soil Conservation Society of America, Ed El-Swaify, S.A., Moldenhauer, W.C., Lo, A.

Wischmeier, W.H. (1976), "Use and Misuse of the Universal Soil Loss Equation" Journal of Soil and Water Conservation, Jan-Feb

Wischmeier, W.H., Smith, D.D. (1978), "Predicting Rainfall Erosion Losses - a Guide to Conservation Planning" Agricultural Handbook No 537, USDA

Zhou, Q., Garner, B.J. (1990), "Pastoral land management in the Australian Arid Zone Using GIS and Remote Sensing Technology" Proc European Conference on Geographical Information Systems, Amsterdam

FEATURE EXTRACTION AND RECOGNITION SYSTEM FOR GIS

A.K. Taib and J.C. Trinder
School of Surveying
University of New South Wales
P.O. Box 1, Kensington, NSW 2033
Australia.

ABSTRACT

Digital image processing is now widely used in extracting useful information from aerial photographs and other images. Conventionally, terrain features are recognized and extracted manually by a human operator viewing imagery on a stereoplotter. The identification of terrain features is based on the interpretation of the characteristics of the image such as tone, size, shape, texture, shadow, pattern, site and interrelationship between features. This paper will describe a system being developed for extracting and recognizing terrain features from analogue and digital images for input into a GIS. The extraction of terrain features by selection and digitizing from the photogrammetric model is done by the operator. The coordinates of digitized points are recorded in the digital file. For feature recognition, this data file must be structured and parameters for geometric analysis such as size and shape of objects must be computed and stored in an attribute file. For texture analysis, digital images are used to compute texture features based on neighbouring gray level dependence matrix (NGLDM). Texture is very important characteristic to identify different types of vegetations and land uses.

INTRODUCTION

Geographic Information System (GIS) provide different services in spatial data manipulation. They allow large volumes of spatial data to be accessed, analyzed and modified through an interactive dialogue with a digital database. However, before one can take full advantage of the functions of using GIS, the data must be acquired and entered into the database. GIS data is input by many different methods such as field survey, existing maps, aerial photographs and satellite imagery. Field survey is slow and very expensive. Existing maps are often out-of-date and suffer from limited and variable geometric accuracy, which can create problems when merging the digitized data into the GIS database. The spatial resolution and geometric accuracy of aerial photographs are higher than that of satellite images. Therefore, in order to obtain an accurate and up-to-date database, a major component of the input of spatial data for GIS must be derived from aerial photographs (Taib et al, 1991).

During the past few decades, photogrammetry has grown tremendously and become a dominant mapping tool all over the world. We are now in the era of digital photogrammetry with its roots in computer technology, computer imaging and analytical photogrammetry (Helava, 1991). Feature extraction from photography requires identifying or recognizing geographic features and assigning feature codes. Up to now, recognition tasks have been performed by skilled operators who have had considerable training and experience, and can recognize features with a high degree of accuracy and efficiency. However, even highly skilled operators do not always work at peak efficiency, and the tedious nature of the tasks can inhibit both accuracy and efficiency. Also, the supply of highly skilled operator is limited, a problem which becomes increasingly severe as the quantity of photographic image data to be processed grows. More importantly, manual processes are expensive and slow.

Data input to GIS must be adequately coded and structured to include the basic GIS data components including the spatial data, the topological data or spatial relationship and the attribute data. Attribute data such as feature names and feature codes can be entered manually and/or interactively, or automatically during or after the spatial data acquisition process provided the features are recognized (Taib et al, 1991). The acquisition of spatial data can be done by using an analytical plotter (AP) or a digital photogrammetric workstation (DPW).

Recent advances in expert system technology offer a means for developing a system for the automatic recognition of geographic features in aerial images (Taib, 1991 and Taib et al, 1991). For feature recognition, parameters of geometry, context and texture may be used as elements in a knowledge base which are the basis for feature recognition using an expert system. A prototype expert system for recognizing geographic features such as bungalows, rural houses, linkhouses, roads, rivers, railways, vegetations and land uses has been developed in a versatile PC-based expert system software Level5 Object (Taib et al, 1992). In the next section, the design issues will be discussed, followed by the processing step in the system, experimental results and discussion , and finally summary and conclusions.

DESIGN ISSUES

Issues which affect the design of a feature extraction and recognition system are concerned with how the data derived from the aerial photographs will be manipulated for the extraction of geographical features and their attributes. They include,which spatial, spectral and contextual attributes are required to adequately define and identify an object in an image; whether any ancillary data is required for feature recognition; which attributes are system derived or user supplied; and how exactly the attributes will be described in the system.

Other issues which need to be considered during the design process of the system include which knowledge representation method will be chosen in the expert system. This is because the recognition of geographic features is very much dependent on the knowledge and ways they are represented in the knowledge base of an expert system. Commonly used knowledge representation methods are rules, semantic networks, logic, frames and object-oriented method. In this research, we proposed to use the object-oriented method because it has many advantages over the other methods. The design and development of knowledge bases using the object-oriented method was given in (Taib et al, 1992).

Many of these issues can only be finalized through initial experiments and prototyping. Prototyping is a low risk design approach where a prototype system can be produced rather quickly, to exhibit all the important features of the final system. Prototyping serves primarily as a tool to refine the problem definition and discover aspects of the problem not visible during the conceptual design of the system. For example, only during the implementation of the prototype in this research did it become apparent which attributes were required or desirable for the recognition of a geographic feature. Many of the attributes used in feature recognition are difficult to describe formally in an automated system, because they cannot be quantified or derived easily. This will result in some attributes not being extractable at all without human intervention, others being extractable only incompletely.

Spectral characteristics of geographic features tend to change, depending on the photographic conditions. However, spatial properties of features such as size(area), shape and texture are very stable and will become the distinguishing characteristics for defining and identifying geographic features (Nagao et al, 1980). For example, residential buildings such as bungalows and linkhouses are recognizable by their size and shape. Therefore size and shape were used as geometric properties to recognize features (objects). Shape can be defined as the outline of an object where it can be computed as the ratio of its squared perimeter to its area.

Texture is an important feature to characterize and discriminate vegetation regions and land uses such as forest, rubber, padi, sundry tree cultivation, grassland, etc. We can find many examples of texture in an aerial photographs: fine texture of low contrast is grassland, coarse texture of high contrast is a forest area while regular texture of large elements is a residential area. There are many different texture analysis methods have been applied to the land use classification problem. Some of the methods are statistically based while others structural approach. A review of both approaches is given in (Haralick, 1979 and Davis, 1981). In statistical approach, texture can be described by tone and its statistical characteristics of tonal variation. Texture features of an image are therefore defined by the overall spatial relationships between the gray tone in the images. On the other hand, the structural approach is based on the model that textures are made of a set of elements arranged according to some placement rules. The statistical approach works well for characterizing natural textures such as in grassland and forest area while the structural approach is more suitable for artificial texture in a residential area which is composed of regularly arranged houses (Nagao et al, 1980). In this research, we have used texture features based on Neighbouring Gray Level Dependence Matrix (NGLDM) developed by Sun (Sun et al, 1983) namely: fineness, coarseness, nonuniformity, second moment and entropy. The major advantages of their approach are as follows : (a). Texture features can be easily computed; (b). They are essentially invariant under spatial rotation; and (c). They are invariant under linear gray level transformation and can be made insensitive to monotonic gray level transformation.

Another important aspect is context or spatial relationships between geographical features which exists in the real world. Context can give important information on the types of objects that can be found or expected in certain areas such as in industrial area, commercial area, urban area, suburban area, residential area or rural area, etc. Contextual information relies heavily on both local and global or world knowledge and experience, which is very difficult to encode efficiently in the knowledge base of an expert system, because the number of possible spatial relationships among geographic features is very large. For geographical features, the spatial relationships may include for example, connectivity of linear features, adjacency of areal features, intersection of features and containments (Taib et al, 1992).

PROCESSING STEP IN THE SYSTEM

"Figure 1" shows the flow diagram of the system for feature extraction and recognition for GIS, where the input data is derived from either analogue and/or digital images. Analogue images are black and white aerial photographs (23cm x 23cm) with 60 percent overlap taken with a wide angle (WA) or superwide angle (SWA) aerial camera. Digital images are produced by scanning the diapositives using a scanner digitizer usually in 8 bits grey scale values for black and white photography. The main processing steps of the feature extraction and recognition system are as follows:

1. Feature extraction

Feature extraction is the process of selecting and delineating geographic features located on the surface of the earth such as buildings, roads, rivers, railways, vegetation regions, etc. Feature extraction can be done either by using an analytical plotter (e.g. ADAM MPS2) on analogue images or a digital photogrammetric workstation (e.g. Leica DPW HAI 710) on the digital images. The ouput of the feature extraction process is a vector data file (spaghetti data file), which consists of the coordinates x,y,z of the digitized points.

2. Data structuring and polygonization

The main aspect of this step is the need to organize the structural relationships between spatial features and to assign appropriate attributes. The structure of polygon adopted for these developments is similar to POLYVRT structure in which a polygon is described as a collection of chains (Suparman, 1992). A chain is defined as a collection of segment lines, which separate two non-overlapping areas. A segment line comprises two node points (start and end point) and additional intermediate points in the case of curved segment line.

The polygons are derived by processing the coordinate data through a logical trace of the sequence of lines defining the boundaries. The detail explanation of the polygonization process is given in (Suparman, 1992). The computer program developed by Suparman was modified to extract the perimeter, area (size) and shape of each closed polygons as geometric attributes, and output to a file (geometric file). This software module was developed in PC environment and written in C Language. The other output files created by the program include polygon, link and node files (structured data files).

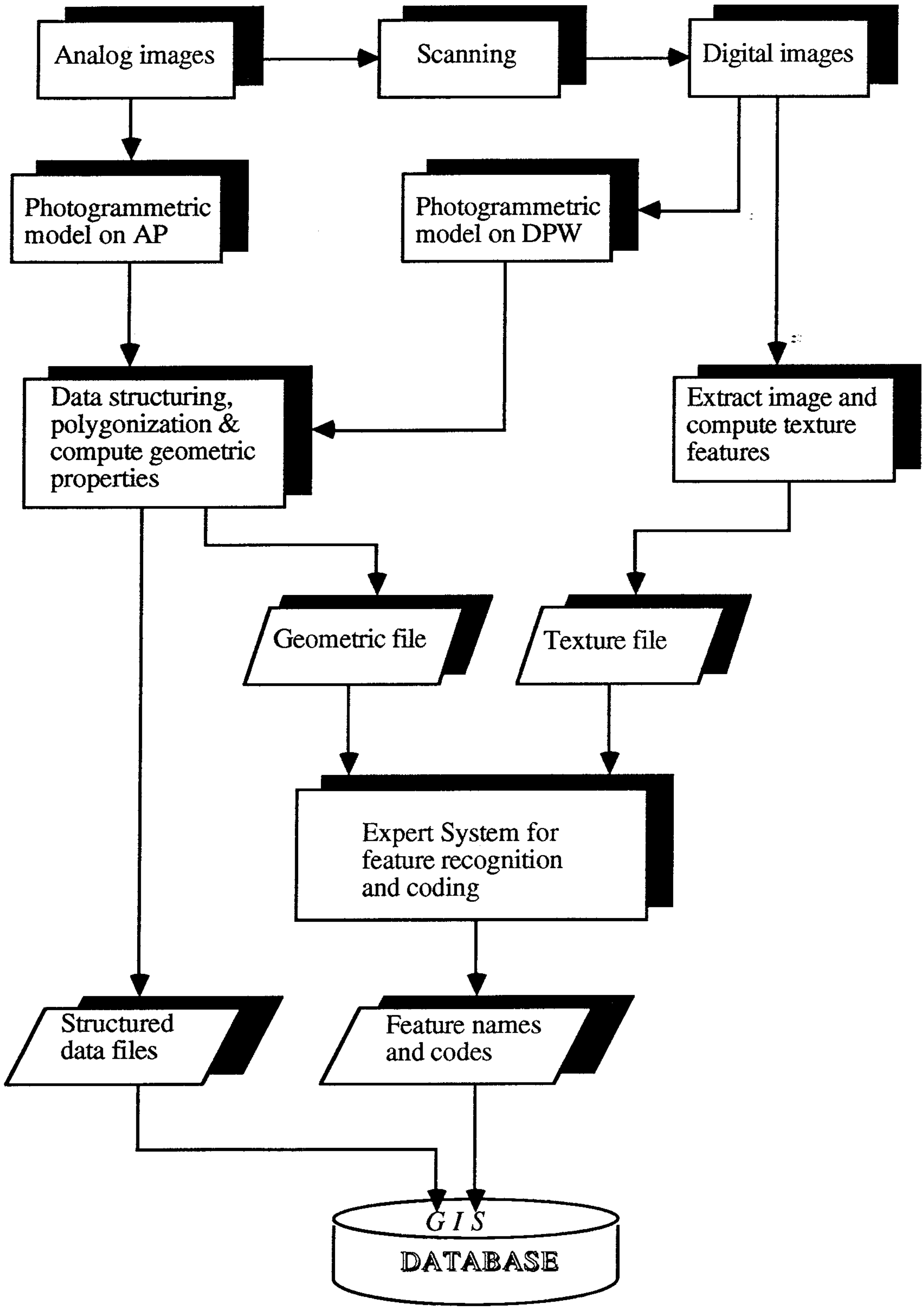


Figure 1 : The flow diagram of the system for feature extraction and recognition for GIS

3. Texture feature extraction

Texture feature extraction is done in two stages using SUN workstation. In the first stage, we extract the representative samples of the known images in the test area, and compute the texture features for the purpose of using them as elements in the knowledge base of the expert system for feature recognition. A computer program was developed to extract digital images and compute texture features based on Sun's method. The detail explanation of the functions of the texture features are given in (Sun et al, 1983). In the second stage, involves the inverse transformation of the model points (in ground coordinates) measured on the analytical plotter (AP) or digital photogrammetric workstation (DPW) to the photo points (in photo coordinates) using the collinearity equations. The photo coordinates are then used to extract digital images from the scanned image file by giving certain window (block) size (e.g. 32x32, 64x64, 128x128 or 256x256 pixels) and compute the texture features as in the first stage. For this purpose, another computer program was developed to perform the tasks in this stage, and the computed texture features are written to a file (texture file) for further processing by the expert system.

4. Expert System for feature recognition

The development of an expert system for recognizing features and determining their feature codes was implemented on a PC-based expert system software called Level5 Object. The first prototype system used only information about geometric properties such as perimeter, area (size) and shape as rules or elements in the knowledge base to recognize geographic features and facts about feature names and codes, and it was not able to recognize many features and different types of vegetations and land uses as indicated by the results of case study I. With the addition of information such as texture and context, it is expected that the system will be able to recognize many more features and different types of vegetation such as forest, rubber, padi, grassland, sundry tree cultivation, etc. Currently, there are only two input files to the expert system namely geometric and texture file as shown in "Figure 1". In future work, the context file will be incorporated into the system. Before these files can be processed by the expert system, they have to be converted into dBASE III format, because the expert system developed in this research can only read the data from files in this format. The output of the expert system is a list of polygon ID with assigned feature names and codes. Finally, the structured data files namely polygons, link and node files from step (3) above, and the output feature names and codes from expert system will become input to the GIS database.

EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we present some of the results of our experiments to study the usefulness of the geometric properties and texture features for recognizing geographic features in suburban area and rural area. The data sets were made available from Malaysia. A brief description of the data sets and the results of the experiments will be presented under the following Case Study I and Case Study II.

Case Study I

The objective of this case study was to evaluate the usefulness of the geometric properties for recognizing geographic features in suburban areas. The area name is Salak Selatan, which located in Kuala Lumpur, Malaysia. The aerial photographs with 60 percent overlap were taken in July 1974, using a wide angle (WA) camera Wild RC10. The scale of the photography is 1: 5000 with the altitude approximately 765 metres above mean sea level. This area covered by vegetations (rubber, forest, grassland), residential buildings (bungalows, rural houses and linkhouses), roads and railway.

The digitizing of the stereomodel was done on the analytical plotter (AP) ADAM MPS2. There were 1667 points and 187 lines digitized in this area. After structuring and polygonization, there were 178 closed polygons and 3 open polygons with 178 centroids and 187 nodes. The geometric parameters such as perimeter, area and shape of each closed polygon were computed and written in a file. This geometric file was then converted into dBASE III format (e.g. dbf file format) and processed by the expert system. The result after recognition by the expert system, 168 out of 178 closed polygons or 94 percents of the closed polygons were correctly recognized. The contigency table is given in Table 1. It shows that by using the geometric rules only, the expert system was not able to recognize correctly some of the features such as railway, rural houses and residential land.

		ASSIGNED FEATURES							
		B	Rh	Lh	Rd	Rw	Vg	Rl	Total
TRUE FEATURES	Bungalow (B)	129	0	0	0	0	0	0	129
	Rural house (Rh)	8	13	0	0	0	0	0	21
	Linkhouse (Lh)	0	0	14	0	0	0	0	14
	Road (Rd)	0	0	0	4	0	0	0	4
	Railway (Rw)	0	0	0	1	0	0	0	1
	Vegetation (Vg)	0	0	0	0	0	7	0	7
	Residential land (Rl)	0	0	0	0	0	1	1	2
	Total	137	13	14	5	0	8	1	178

Table 1 : Contigency table for feature recognition of Salak Selatan Test Area

Case Study II

The main objective of this case study was to evaluate the usefulness of the texture features to characterize the different types of vegetations and land uses. Before we can use the texture features as elements in the knowledge base of the expert system for feature recognition, we have to analyse the texture features of the representative samples extracted from the known images of the test area. This process was considered as the training phase (first stage) before the second stage and recognition phase by the expert system. This case study will only give the results of the training phase because the data has not yet been processed by the expert system.

The test area for this study is a rural area in the Northern State of Malaysia namely Jeli. The aerial photographs with 60 percent overlap were taken in November 1978, using a wide angle (WA) camera Wild RC10. The scale of the photography is 1: 25000 with the flying height approximately 3822 metres above mean sea level. This area mostly covered by forest, rubber, padi, grassland, sundry tree cultivation, roads, river, village (rural residential) and town area.

The scanning of the diapositives was done using the Helava DSW 100 Digital Scanning Workstation, at the Queensland University of Technology, Brisbane. The DSW 100 consists of a precise scanner, which is specially designed for high quality image scanning. Pixel size or resolution can be selected between the range 8 - 165 μm and image scanning rates are approximately 40000 pixels per second (Miller et al, 1992). For this study, the pixel size 13.4

μm (nominally 12.5 μm) with 8 bits grey scale was selected. For 60 percent overlap, each photograph required 235.30 MBytes memory.

To process this digital image, image processing algorithms were developed as explained in previous section, which are capable of extracting relevant information from the image and computing texture features, that characterize different types of vegetations and land uses. The block size of the sample selected are 32x32, 64x64, 128x128 and 256x256 pixels. The experiments were carried out for both left and right images of Jeli, and for every feature at least 10 - 15 samples per block size were extracted for texture analysis. Results given in this paper indicate the significance of some textural features such as second moment and entropy for describing the characteristic of the image. These two texture features are the measures of the homogeneity and coarseness of the image respectively. The higher the value of the second moment, the more homogeneous the image; the smaller the value of the entropy the coarser the image.

Plots of texture of geographic features are ordered according to the complexity of the image in Figures 2 and 3, which show the range of values for the second moment feature of the representative geographic features namely padi field, vacant land, rubber area, sundry tree cultivation, forest area and residential area. Figures 4 and 5 show plots of the range of values for the entropy of the same geographic features. These texture features can be used as additional rules in the knowledge base of the expert system to recognize padi, vacant land, rubber, sundry tree cultivation, forest and residential area. Results in Figures 2,3,4 and 5 show that second moment and entropy can be used to separate between padi, rubber, forest and residential area, but are not sufficient to separate vacant land from rubber, or sundry tree cultivation from forest or rubber. These problems can only be resolved in the expert system by using other information such as geometric properties and contextual information to separate and recognize them. However, much more experimentation and analysis should be done for the texture features to include other types of vegetations and land uses.

SUMMARY AND CONCLUSIONS

This paper has described the system for feature extraction and recognition for GIS. Feature extraction is performed manually by the operator in the analytical plotter (AP) or digital photogrammetric workstation (DPW) by selecting and delineating the features in the stereomodel. Feature recognition tasks are done automatically by an expert system. Regularly shaped objects such as residential buildings (e.g. bungalows and linkhouses), are usually very easy to recognize by their geometric properties, but for irregular shape objects or natural features such as vegetation areas, texture features are important for identifying the different types of vegetations and land uses. To resolve ambiguities, contextual information (e.g. adjacent, containment, etc) should therefore be used in feature recognition process, especially those features likely to be found in urban, suburban, residential, commercial and industrial areas such as different types of buildings, railway features and road features.

The results obtained from the two case studies are encouraging, and future work will involve refining the prototype expert system, as well as increasing the complexity by adding the textural and contextual information in the knowledge base of the expert system for feature recognition, and the analysis of texture features for other types of vegetation and land use by using digital images at different image scales and pixel sizes. Furthermore, our ultimate goal in this research is to recognize geographic features automatically from digital images, and to achieve this goal we hope in the future the feature extraction process can be done automatically using digital image processing techniques which includes image enhancement and segmentation techniques, in the digital photogrammetric workstation (DPW).

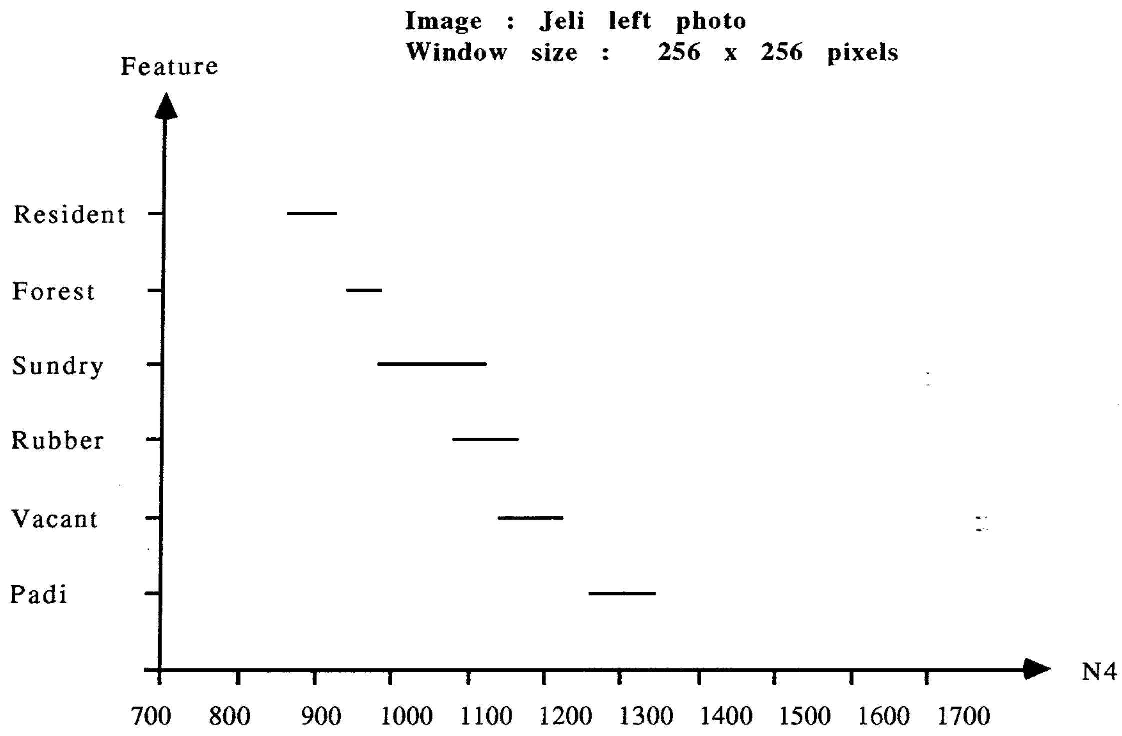


Figure 2 : Range plots for second moment, N4.

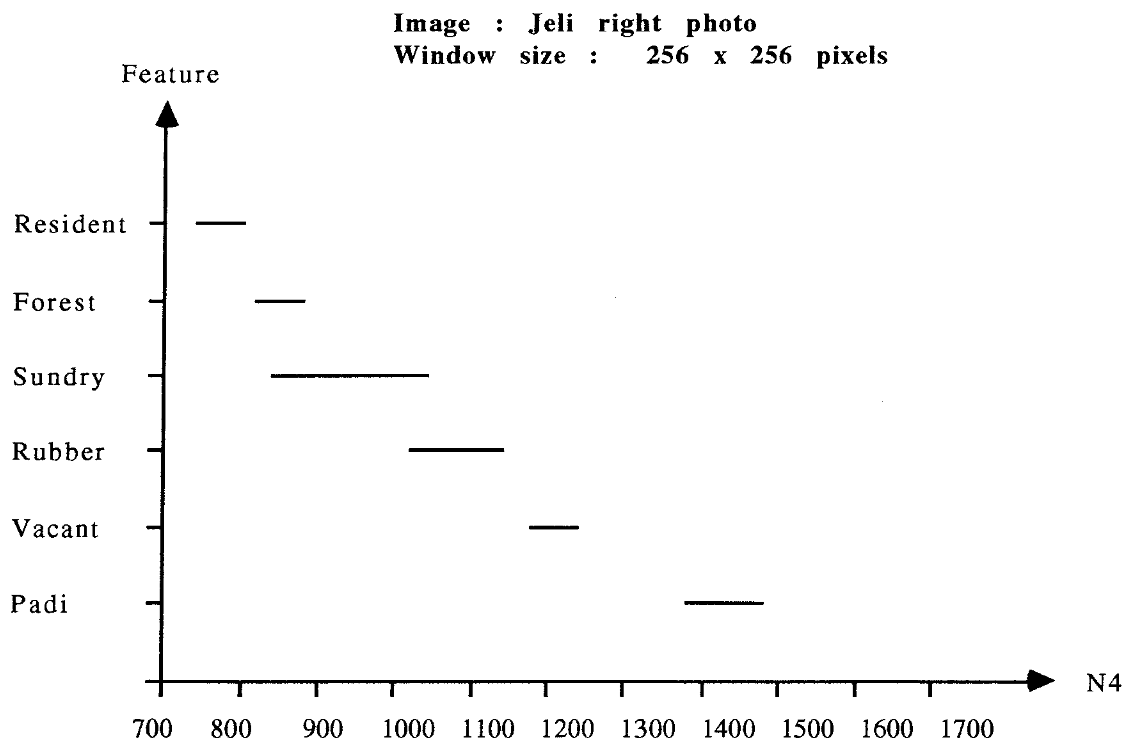


Figure 3 : Range plots for second moment, N4.

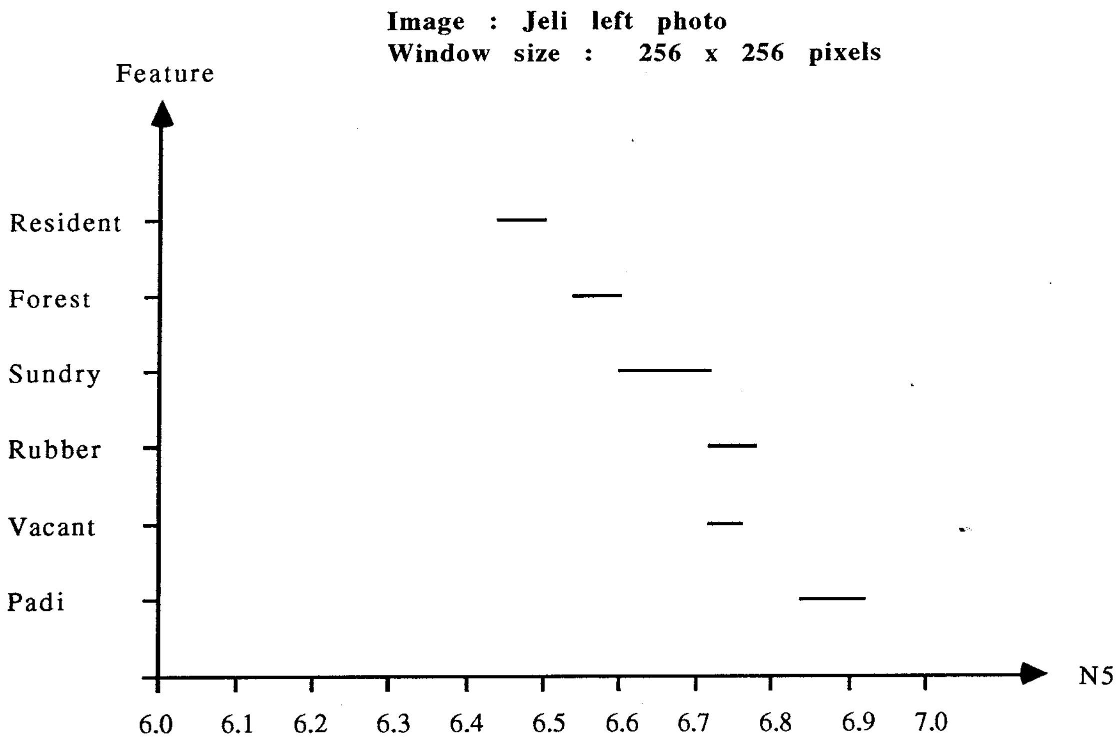


Figure 4 : Range plots for entropy, N5.

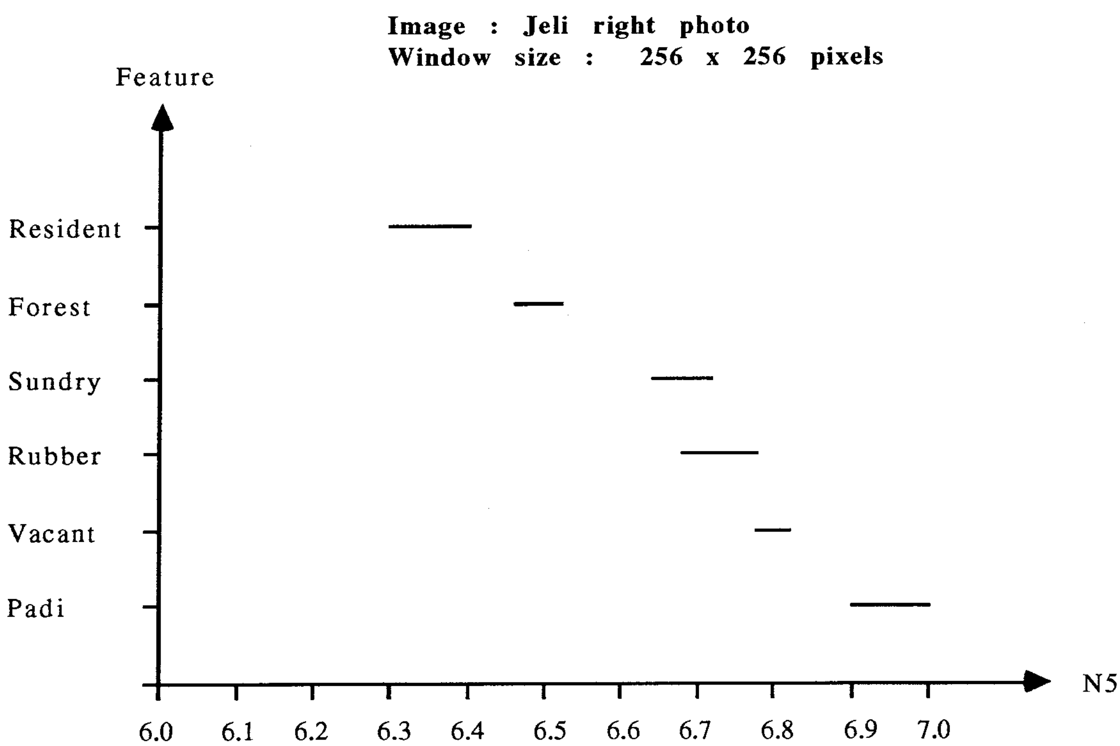


Figure 5 : Range plots for entropy, N5.

ACKNOWLEDGEMENTS

The authors thank the Director General of Survey and Mapping Malaysia for providing the aerial photographs and other related materials for this study. We would also like to thank Mr. Maurice Friend at the Queensland University of Technology, Brisbane, who helped in scanning the aerial photographs of Jeli by using the Helava Scanner DSW 100.

REFERENCES

Davis, L.S. (1981), "Image Texture Analysis Techniques - A Survey", Digital Image Processing, Edited by J.C. Simon and R.M. Haralick, D. Reidel Publishing Company, Dordrecht, Holland 1981.

Haralick, R.M. (1979), "Statistical and Structural Approaches to Texture", Proceedings IEEE 67, 1979.

Helava, U.V. (1991), "Prospects in Digital Photogrammetry", The Photogrammetric Journal of Finland, Vol. 12, No. 2, pp. 57-64.

Miller, S.B. and Thiede, J.E. (1992), "A Line of High Performance Digital Photogrammetric Workstations - The Synergy of General Dynamics, Helava Associates, and Leica", International Archives of Photogrammetry and Remote Sensing, Vol. XXIX, Part B2, Comm. II, pp. 87-94.

Nagao, M. and Matsuyama, T. (1980), A Structural Analysis of Complex Aerial Photographs, Plenum Press, New York.

Sun, C. and Wee, W.G. (1983), "Neighboring Gray Level Dependence Matrix for Texture Classification", Computer Vision, Graphic and Image Processing 23, pp. 341-352.

Suparman, A. (1992), An Incremental Approach to the Implementation of Land Information System in Indonesia, Unpublished Ph.D. Thesis, The School of Surveying, UNSW.

Taib, A.K. (1991), "Expert Systems for Data Collection by Photogrammetry", Proceedings First Australian Photogrammetric Conference, 7-9 November, 1991, UNSW, Sydney, Australia, Vol. 2, Paper No. 35.

Taib, A.K. and Trinder, J.C. (1991), "LIS/GIS Data Acquisition by Photogrammetry", AURISA 91 Proceedings of the Nineteenth Australasian Conference on Urban and Regional Information System, Wellington, New Zealand, 19-22 November, 1991, pp. 362-371.

Taib, A.K. and Trinder, J.C. (1992), "An Object-Oriented Approach to the Design of Knowledge Bases and Geographic Information Systems", AURISA 92 Proceedings of the 20th. Australasian Conference on Urban and Regianal Information System, Gold Coast, Queensland, 25-27 November, 1992, pp. 149-156.

GEOGRAPHICAL INFORMATION SYSTEMS AND LAND-USE/TRANSPORT INTERACTION MODELS : TOWARDS SYSTEM INTEGRATION

T T Ton and J A Black
Department of Transport Engineering
School of Civil Engineering
University of New South Wales
PO Box 1, Kensington NSW 2033
Australia

ABSTRACT

The issue of integrating GIS and land-use/transport interaction (LUTI) models has become one of main themes in applied research projects in some countries. However, details on specific techniques (or models) or precise descriptions of any difficulties in using the systems are rarely reported in the literature. This paper identifies the need for integrating GIS and LUTI models and reviews literature on this experience. The question of how GIS and LUTI models should best be integrated is then addressed with particular reference to the definition of GIS, its role in the integration process, and the perception of the intended developer/user toward the integrated system. Possible developmental configurations for integrating GIS and LUTI models are discussed from two different viewpoints: developer and user. Major software implementation strategies - development from scratch, from a library of routines, or from off-the-shelf products - are discussed. A case study for integrating GIS and LUTI models is described in the format of the system description (objective, GIS and LUTI models integration configuration, implementation approach), input and output data. A case study of research in progress is presented. The aim is to develop a computer-based spatial information system encoded for traffic impact analysis that will allow transport planners to explore the implications of alternative land-use/transport strategies.

INTRODUCTION

The application of GIS to transportation analysis has focused primarily on inventory, or descriptive-type, problems such as pavement management, accident inventory and analyis, and maintenance management (Vonderohe, 1992). In Australia, for example, AUSROADS (the Association of Australian Road Authorities) have displayed national freight movement data graphically using GIS (Turner, 1991). Attention has recently turned to identifying the role of GIS in traditional transportation, transit, and traffic planning and analysis (O'Neill, 1991). The issue of integrating the two systems together (GIS and LUTI Systems) has become one of main themes in applied research projects in many countries (Patterson, 1990; and Bates and Schwetz, 1990). However, details on specific techniques (or models) or precise descriptions of difficulties in using the systems, are sometimes scarce (Lewis, 1990).

This paper describes a framework for integrating GIS and LUTI models. The next section identifies the need for integrating GIS and LUTI models. Section three presents a summary of the literature on this experience. In section four, the question of how GIS and LUTI models should best be integrated is addressed with particular reference to the definition of GIS, its role in the integeration process, and the perception of the intended developer and user towards the integrated system. Major possible software implementation strategies - development from scratch, from a library of routines, or from off-the-shelf products - are discussed in section five. Section six presents preliminary ideas on a GIS and LUTI models integration project, which is currently part of collaborative research between the Roads and Traffic Authority New South Wales and the University of New South Wales.

This collaborative project aims at developing a computer-based spatial information system encoded for traffic impact analysis that will allow transport planners to explore the economic, social and environmental implications of alternative land-use/transport strategies. It builds on existing technologies of land information systems, database mangement systems, land-use/transport modelling, environmental impact modelling and GIS. With the aid of the GIS subsystem, different transport planning scenarios can be evaluated graphically with the following measures: land use information; transport network information; travel demand information; travel performance; road safety; energy performance; environmental performance; equity/social measures; and economic performance.

INTEGRATING GIS AND LUTI MODELS: THE NEED

The acquisition of data, the management of data, the selection and the use of appropriate analytical tools that suit an overall objectives within a specified constraints (budget, data availability, modelling preference, etc.) and the assessment of alternative land-use/transport solutions are core activities in the planning of economic and environmentally sensitive transport systems. In general, data acquisition is undertaken manually and the management of data is very poor (eg. data exist in flat files and are not incorporated under any database management system, or they are scattered across different systems/organisations from which they are difficult to integrate). Consequently, data acquisition and data management incur a significant part of the total expenses involved in the analytical process.

The transport analytical tools are mostly implemented in a form of a computerised four-step transport planning model (trip generation, trip distribution, mode choice and traffic assignment) with exogeneous inputs of land use. This system, referred to as land-use/transport interaction (LUTI) models, is used to generate implications of a number of planning policies. In contrast to the other planning activities, the development of the analytical tools - both theoretical and computational - has been attracting more interests and/or investments. For instance, there are many commercially available LUTI packages which implement similar models. Details of the main features of the most popular packages used in USA are listed in Urban Transport Monitor (1992a,b) and those used in Australia are summarised in Black (et. al., 1993). The reasons for such investment are increasing demands for better performance of the packages themselves (in terms of a more flexible modelling environment) and their integration with other models/systems - graphic user interface, geographical information system, knowledge-based expert system, environmental impact models, economic evaluation models, road safety models, and energy models.

The assessment of alternative land-use/transport plans rely heavily on the quality of data acquisition, data management, and the outcome from the selection and the use of the software package. If the acquisition of data is automatic and the database management system offers a reasonable cost to extract data, if the LUTI model system offers a flexible and integrable system, and if its application of LUTI models to a particular planning study is well-structured and well-interpreted, then the assessment of alternative land-use/transport solutions in terms of economic and environmentally sensitivity can be accomplished. However, the assessment procedures of alternative land-use/transport options can be restrained by costly data acquisition and inefficient data management.

These problems of the transport planning process are practical limitations that a geographical information system (GIS) attempts to address. A geographical information system (GIS) is a computerised database management system (DBMS) for the capture, storage, retrieval, analysis, and display of spatial data. With this definition, a structure of GIS generally consists of a graphical user interface (GUI) (ie. used for displaying data), a DBMS (ie. used for data storage and retrieval) and a spatial modelling function library (ie. used for capture and analysis of spatial data). More details on approaches to defining GIS can be found in Cowen (1990) and Dangermond (1990).

INTEGRATING GIS AND LUTI MODELS: LITERATURE REVIEW

A literature survey focused on latest information relating to the integration GIS and transport modelling. In Australia, there is little evidence of reports on the result of GIS and LUTI model integration. A joint project by Roy (from Australia) and Snickars (from Sweden) reported on the CARP (Computer Aided Regional Planning) model (Roy and Snickars, 1992). CARP is an integrated land-use and GIS system designed as an interactive environment for the generation and evaluation of land-use plan alternatives. Other examples are Black and Ton (1989); and Davison (1992) of VICROADS, where, in both cases, computer graphic routines were developed to enhance the presentation of land-use/transport planning results.

From the overseas literature, there are a number of key papers which, to some extent, could be seen as collective representations of the GIS and LUTI integration. These are: Wegener and Junius (1991) from Europe; Nakamura and Shimizu, (1990) from Japan; and Vonderohe (1992a,b) from the United States. There are large gaps between the diffusion of geographical information systems in European countries. Whereas in some countries, notably in the United Kingdom and the Netherlands, GIS technology has already become standard practice in cartography, local land management and planning, other countries, such as Germany, have been slow in adopting the new technology. Wegener's hypothesis is that national traditions determine the institutional environment and procedures into which the new technology needs to be integrated. This restricts the transfer of 'universal' GIS software, by which Wegener means GIS shells such as ARC/INFO of ESRI Environmental Systems Research Institute (Redlands, CA), which have been developed in different national settings.

In Japan, GIS is widely applied to a variety of urban management tasks covering city planning, utility network management and automobile navigation. It appears that the Japanese have placed a strong focus on the application of GIS to automobile navigation. The general level of applying GIS to transportation in many US State Department of

Transportation (DOT), has a similar pattern to that of Europe. Whereas some DOTs such as Wisconsin and Ohio, have already applied GIS in many planning activities, other state DOTs are still slow in adopting this new technology (Vonderohe, 1992a,b).

What can we learn from all of these experiences? The following quote from Vonderohe's paper is worth considering, and reflects one perspective.

> "Principal aspects of the institutional context for GIS adoption and application by DOTs include issues of determining the most critical applications that must carry the brunt of initial GIS base map data acquisition costs, sharing costs across applications, gaining and retaining support of high level management and of the public, coordinating with other state agencies and with external organisations, and utilising standard developments." (Vonderohe, 1992b, p.42).

Before presenting our case study, we discuss alternative developmental configurations and software implementation strategies.

INTEGRATING GIS AND LUTI MODELS: DEVELOPMENTAL CONFIGURATIONS

A key point for the discussion is that GIS and LUTI models have been developed as two independent systems. Should GIS be part of LUTI system? or should the LUTI system be part of GIS? One could speculate that there would be two developmental trends: LUTI packages are extended with capabilities to cover GIS-type functions; GIS is extented for use in transport modelling. The literature review confirms the existence of both of these trends, but there are others. Several major LUTI-type package developers have been reviewing the provision of links to GIS (Lewis, 1990, p.37). Some are enhacing their transport network editing and geographical display capabilities (Vonderohe, 1992b, p.39; and Urban Transport Monitor, 1992b, p.11). TRANSCAD, developed by Caliper Corp, is a typical example of a LUTI package with GIS capability (Urban Transport Monitor, 1992a,b, p.11). Commercially-available GIS packages have been extended for use in transport modelling, but are also useable for a large number of non-transportation GIS applications (Vonderohe, 1992b, p.39).

The question of how GIS and LUTI models should best be integrated is now addressed. The authors argue that there is no one technical basis for incorporating GIS and LUTI models. It depends on the definition of GIS, its role in the integration, and the perception of the intended developer (and user) toward the integrated system. Possible developmental configurations for integrating GIS and LUTI models are indicated in Figure 1, which shows the six conceptual models of integration. In Model One only a full-scale GIS is used. In Model Two both a full-scale GIS and a full-scale LUTI system are used, but these are not linked together. In Model Three both a full-scale GIS and a full-scale LUTI systems are used, but these are linked together through an interface. In Model Four a full-scale LUTI system is extended to cover a partial GIS. This configuration represents the first trend which was discussed above from the LUTI system developer's perspective. In Model Five a full-scale GIS is extended to cover a partial LUTI system. This configuration represents the second trend which was discussed above from the GIS developer's perspective. Finally, in Model Six a full-scale GIS is integrated with a full-scale LUTI system.

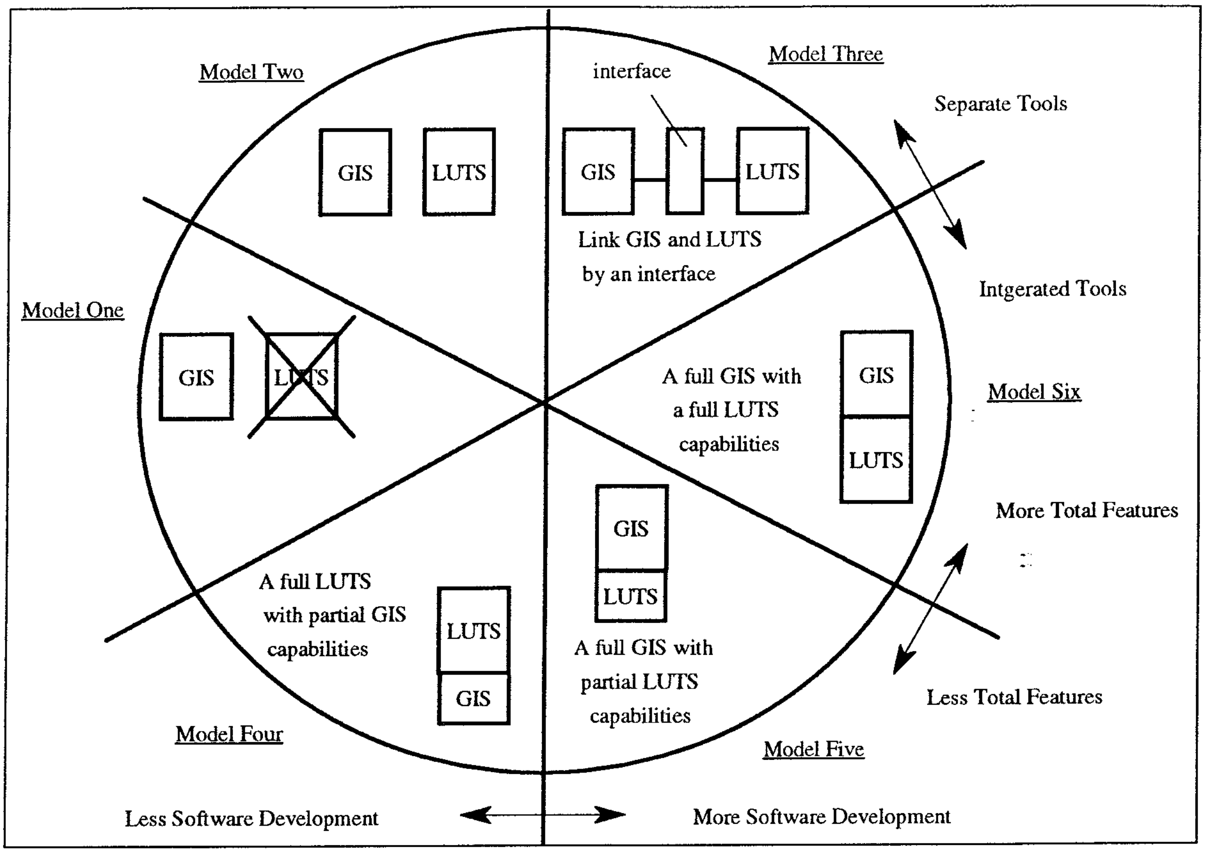


Figure 1: Possible Developmental Configurations for Integrating GIS and LUTI System (adopted from Lewis, 1990)

The definition of GIS, its role in the integration, and the perception of the intended developer or user toward the integrated system are generally reflected from the viewpoints of the software's developer and software's user. Table 1 can be used as a decision matrix which associates each model with a set of criteria from both software developers (the choice between more or less software development) and the users' viewpoints (the choice between more or less total features, and the choice between integrated or separate tools). Selecting a developmental configuration for integrating GIS and LUTI system from Table 1 is a very important decision as it represents the balance between software development and the features and tools offered by the system to the users.

Model	Software Development		Total Features		Tools	
	Less	More	Less	More	Separate	Integrated
1	Yes	-	Yes	-	Yes	-
2	Yes	-	-	Yes	Yes	-
3	-	Yes	-	Yes	Yes	-
4	Yes	-	Yes	-	-	Yes
5	-	Yes	Yes	-	-	Yes
6	-	Yes	-	Yes	-	Yes

Table 1: Decision Matrix for Integrating GIS and LUTI System

INTEGRATING GIS AND LUTI MODELS: SOFTWARE IMPLEMENTATION STRATEGIES

Selecting an appropriate tool for the job is a key in searching for a successful implementation strategy for integrating GIS and LUTI system. It should be noted that different tools may be needed for different phases of the project. This section discusses the main approaches used in the implementation of GIS, LUTI system, the interface (if any) and the integrated system. Table 2 summarises five approaches which are combinations of the three basic tools in software development: programming language; programming library; and off-the-self product (or programming shell).

Strategy	Description	Example
1	use programming language	C or C++ or PASCAL or FORTRAN
2	use programming language and domain library	C or C++ or PASCAL or FORTRAN and ONTOS C++ library of Ontologic (Burlington, Massachusetts)
3	use programming shell	ARC/INFO of ESRI (Redlands, CA)
4	use programming shell and programming language	ARC/INFO and C or C++ or PASCAL or FORTRAN
5	use programming language, domain library and programming shell	C or C++ or PASCAL or FORTRAN TRANSOOP of Ton and Black (1993) and ARC/INFO

Table 2: Software Implementation Strategies for Integrating GIS and LUTI System

The first strategy has the following advantages: a high level of control and flexibility; full access to hardware and the operating system; open-ended functionality (the programmers are only limited by their own code); open-ended user interface. The disadvantage of this strategy includes all of the risks involved in a development cycle, such as a costly, under-developed, product. Typical example of this strategy is the work by Roberts (et.al., 1991); and Davison (1992). Roberts, et.al. used an object-oriented programming language to develop GeoScope which is an object-oriented GIS that runs on IBM PC on Microsoft Windows. Davison (1992) used FORTRAN to develop GIS-typed graphic (grid-based format) for presenting LUTI system output.

The advantage and disadvantage of using strategy number 2 is similar to using strategy number 1, except the development time can be shorter with the availability of the library which is trade-off for slower runtime. Smith (1991) used ONTOS C++ library to implement a geographic object-oriented databases.

The strength of using strategy number 3 is to gain immediate access to all aspects of the shell. For example, if ARC/INFO is used as a GIS shell then there is an economy in directly using the graphical user interface, the database management system and the spatial modelling function library offered by ARC/INFO. Furthermore, development time might be quicker with the shell's supported language (macro language) without requiring a high level of expertise in programming language. The disadvantage of this strategy is a lower level of control and flexibility; limited access to hardware and the operating system; generally, a lack of open-ended functionality; and no open-ended user interface (if compared with strategy 1 and 2). Furthermore, it is more time consuming to train staff to familiarise themselves with the shell's command than to customise a system which is designed for transport users. Typical examples of using this strategy are discussed in Wegener and Junius (1991) and in Vonderohe (1992a,b).

Strategy number 4 is an extension of strategy number 3 with the use of a programming language. The clear advantage in using this strategy is the system would be able to reach out (via the use of programming language) to more flexible, and user-defined modelling functions. Apart from that, this strategy has all the advantages and disadvantages of strategy number 3. Examples of using this strategy are cited under strategy number 3.

Strategy number 5 is an extension of strategy number 4 with the use of a programming library. The clear advantage in using this strategy is the system would be able to reach out (via the use of programming language) to a more flexible and user-defined modelling library. Rather than a using a programming language and a programming library (strategy number 2) or using the programming shell and programming language (strategy number 4), this strategy combines these strategies together. If the best features of both strategies are implemented then it would be the most efficient strategy otherwise it will suffer from inheriting all the disadvantages from these two strategies. A typical example of using this strategy is an implementation framework proposed for the case study discussed in the next section. A programming shell (eg. ARC/INFO or GENASYS for the GIS), a programming language (eg. C++) and a programming library (eg. TRANSOOP is a C++ language library implementing the land-use/transport interaction developed by Ton and Black (1993)) forms an implementation framework for the case study.

The answer to the question on which software implementation strategy is the most suitable depends on the nature of the application (ie. reflect on the selected developmental configuration), the pressure of immediate implementation, cost-benefit analysis, the current state of the technology, and the anticipated near-future improvements in the technology described by software vendors. Some preliminary thoughts on software implementation strategies are pursued through a case study.

INTEGRATING GIS AND LUTI MODELS: A CASE STUDY

The case study described in this section is based on some preliminary ideas emerging from a collaborative research project between University of New South Wales (Department of Transport Engineering, School of Civil Engineering, and the School of Surveying) and the Roads and Traffic Authority, New South Wales. When a prototype is completed, an integrated GIS and LUTI modelling system could be used in the assessment of alternative land-use/transport plans. The information associated with a particular land-use/transport system design would include land-use information, road network information, travel demand information, economic evaluation, travel performance, environmental impact, safety, equity/social and energy measures (Figure 2). In this way, the detailed analysis of road infrastructure is seen within a proper context of land-use, transport and the environment.

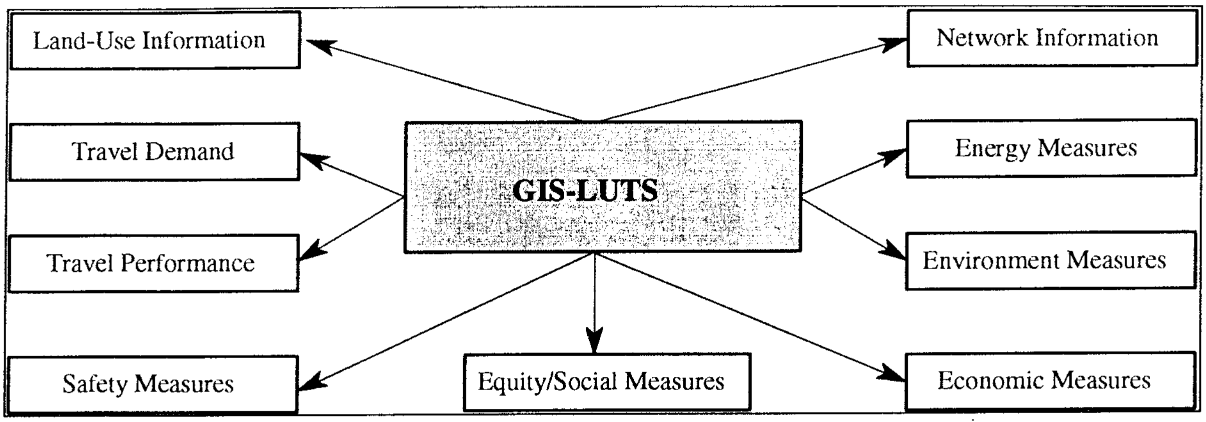


Figure 2: Performance Measure Modules of the GIS-LUTS Integrated System

Underpinning the land-use/transport system (LUTS) is a four-step transport planning model of trip generation, trip distribution, mode choice and traffic assignment. Trip generation models an estimate of the number of trips generated by each basic spatial unit designated for the analytical purposes (zone). The data required are of a spatially-based socio-economic kind, such as population, and employment. Trips generated from each zone are then allocated to other zones to form an origin-destination (OD) matrix using a trip distribution model. Data requirements are the transport impedance matrix (distance, or travel time or generalised cost between zones) and travel behaviour. Modal split models estimate the proportion of trips amongst the competing transport modes between zones. Data required for these models vary considerably. However, they all essentially contain the unit costs associated with each transport mode (eg. parking cost, vehicle operating cost, travel time cost, public transport fare, and access and waiting times). Traffic assignment loads trips for different transport modes onto the networks. The direct result is the estimated traffic flows on every link of each transport network, including the road network. Basic data required by assignment model are the transport impedance matrix and the OD matrix. A number of specialised impact models, such as vehicle emissions, energy consumption, noise pollution, delays and accident risks are then attached to the output of the traffic assignment phase.

The potential role of GIS is to set up inventories of socio-economic data, land use and transport system facilities (eg. the coordinates of road network and associated information such as the geometric standards, history of traffic flows and accidents) Other additional data required, for instance, to correct the basic traffic noise level estimation are ground conditions, distance between the edge of nearside carriageway and the facade. GIS can also be used to in convert data to and from the following three graphic formats for input and output data: linked-based, grid-based and zonal based information. Table 3 compiles performance measures produced by this system. Table 3 adopts (with some modifications) the Roads and Traffic Authority New South Wales performance measures for sketch planning in New South Wales which were reported by Kilsby (et.al, 1992, Table 2, p.206). The appropriate graphic format has been added to this table.

Modules	Performance Measures	Graphic Formats		
		Link-Based	Grid-Based	Zonal Based
Land-use Information	Population (Million)	-	Yes	Yes
	Employment (Million)	-	Yes	Yes
Network Information	Road network size (Lane km)	Yes	Yes	Yes
	Car Travel (Million vehicle km/year)	Yes	Yes	Yes
	Car trips (Million person trips/year)	Yes	Yes	Yes
	Total trips (Million person trips/year)	Yes	Yes	Yes
	Car Trip Distance (Million person km/year)	Yes	Yes	Yes
	Person km/head (000 km/year)	Yes	Yes	Yes
Travel Demand	Origin-Destination Matrix	-	Yes	Yes
Travel Perfromance	Total Vehicle Kilometres	Yes	Yes	Yes
	Local Vehicle Kilometres	Yes	Yes	Yes
	Arterial Vehicle Kilometres	Yes	Yes	Yes
	Freeway Vehicle Kilometres.	Yes	Yes	Yes
	Local Vehicle Hours	Yes	Yes	Yes
	Arterial Vehicle Hours	Yes	Yes	Yes
	Freeway Vehicle Hours	Yes	Yes	Yes
	Average Speed (Km/hour)	Yes	Yes	Yes
	Volume Capacity Ratio	Yes	Yes	Yes
	Spare Capacity (Vehicles/hour)	Yes	Yes	Yes
Economic Evaluation	Road Infrastructure costs ($Million)	Yes	Yes	Yes
	Externality costs-congestion ($Million/year)	Yes	Yes	Yes
	Externality costs - noise ($Million/year)	Yes	Yes	Yes
	Externality costs - air pollution ($Million/year)	Yes	Yes	Yes
	Road externality costs - accidents ($Million/year)	Yes	Yes	Yes
Road Safety	Fatal Accidents (accidents/year)	Yes	Yes	Yes
	Personal Injury Accidents (accidents/year)	Yes	Yes	Yes
	Property-Damaged Only Accidents (accidents/year)	Yes	Yes	Yes
Energy	Total Fuel Consumption (Million litres/year)	Yes	Yes	Yes
	Average Fuel Consumption	Yes	Yes	Yes
Environment Impact	Air Pollution			
	CO2 Emittants (000 tonnes/year) by Road	Yes	-	Yes
	CO Emittants (000 tonnes/year) by Road	Yes	-	Yes
	NOx Emittants (000 tonnes/year) by Road	Yes	-	Yes
	HC Emittants (000 tonnes/year) by Road	Yes	-	Yes
	CH4 Emittants (000 tonnes/year) by Road	Yes	-	Yes
	Noise Pollution			
	Combined Noise Level (Decibel)	Yes	Yes	Yes
Equity/Social Measure	Self containment to/from (per cent)	-	-	Yes
	Car peak share to/from (per cent)	-	-	Yes
	Av peak car time to/from (minute)	-	-	Yes
	Av peak car distance to/from (km)	-	-	Yes
	Av peak trip length to/from (km)	-	-	Yes
	Av peak car trip cost excluding travel time ($)	-	-	Yes
	Av peak car trip cost including travel time ($)	-	-	Yes
	Public Transport peak share to/from (per cent)	-	-	Yes
	Av peak public transport time to/from (minute)	-	-	Yes
	Av peak public transport distance to/from (km)	-	-	Yes
	Av peak public transport fare to/from ($)	-	-	Yes
	Av peak public transport cost excluding travel time ($)	-	-	Yes
	Av peak public transportcost including travel time ($)	-	-	Yes

Table 3: Performance Measures of the GIS-LUTS Integrated System

Although consultative work has not advanced to the stage to reach a determination on the most appropriate way forward, the configuration of Model 3 discussed above in Figure 1 is proposed to structure the system architecture for integrating GIS and LUTS. Figure 3 links GIS and LUTS through a user interface. Furthermore, the proposal employs strategy number five above for implementing the system. It employs a GIS shell (which could be ARC/INFO or GENASYS) and a programming language called C++ to develop the interface linking GIS and LUTS. The development of the object-oriented library of modules on land-use and transport interaction is a further direction in which this research project can go, as explained by Ton and Black (1993).

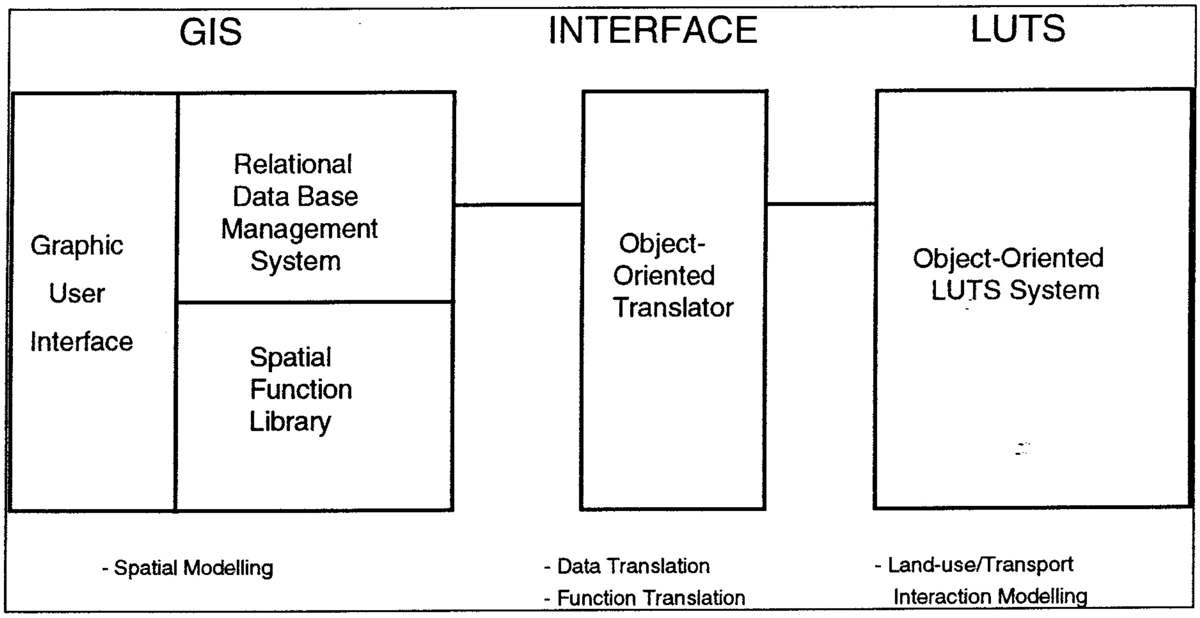


Figure 3: GIS-LUTS Integrated System's Architecture

Object-oriented approaches are important for the development of the interface, whereas the GIS shell only support primitive objects such as points, vectors, lines and polygons. LUTI modelling is more abstract, or contains more composite forms, such as a property boundaries, roads, or houses. The object-oriented interface is designed to translate data and function to and from LUTS and the GIS. Figure 4 illustrates a class hierarchy for translating a composite spatial object into GIS-supported objects.

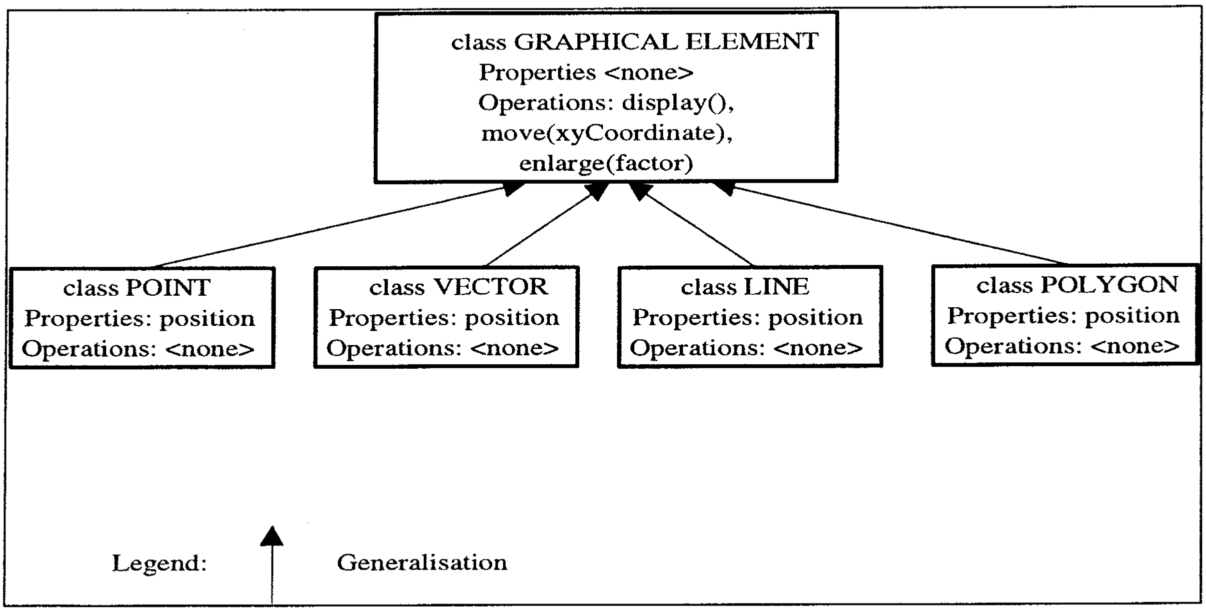


Figure 4: Class Hierarchy for Geometric View in an Object-Oriented GIS (adapted from Roberts, et. al., 1992)

There are four reasons why the object-oriented programing approach is employed in implementing the interface. Firstly, to minimise the risk in the software development by not having to rely heavily on the GIS shell. Secondly, to make a direct link with the LUTS which has been built by the object-oriented technology (Ton and Black, 1993). Thirdly, to improve the economy of the system by using the object-oriented technique to customise the process of calling GIS functions. Finally, the application and testing of this new technology to the enhancement and/or development of GIS in Australia (Smith, 1992) and overseas (Roberts, et.al, 1992) has shown that the object-oriented programming approach is a promising direction for GIS software implementation. Furthermore, object-oriented approaches are beginning to find favour in the simulation of traffic management strategies and the provision of real-time information to drivers - for example, the PROROAD prototype under development by Daimler-Benz AG, Stuttgart (König and Langbien, 1993).

More technical details on the general framework for the GIS development is provided by Rhind (et.al , 1988). Six classes of GIS functions are identified: data input, data manipulation, data retrieval, data analysis, data display, and database management. The object-oriented framework which implement these six basic GIS classes could serve as a generic and flexible framework for developing the interface with GIS. Figure 5 shows the structure of class "Interface" which encapsulates the whole object-oriented framework with six objects of major classes mentioned above. The design of the class "Interface" offers a flexible structure for interfacing between GIS and LUTS. It is flexible because in an instance of class "Interface", one can set one, or all, of the objects of the six basic classes for accessing to/from a GIS application.

```
class INTERFACE

Properties:
        DataInput class: Input_Object
        DataManipulation class: Manipulate_Object
        DataRetrieval class: Retrieve_Object
        DataAnalysis class: Analyse_Object
        DataDisplay class: Display_Object
        DataManagement class: DBMS_Object
Operations:
        INTERFACE() //to create instance of class INTERFACE
        SetDataInput(DataInput class: newInput)
        SetDataManipulate(DataManipulation class: newManipulate)
        SetDataRetrieve(DataRetrieve class: newRetrieve)
        SetDataAnalyse(DataAnalysis class: newAnalysis)
        SetDataDisplay(DataDisplay class: newDisplay)
        SetDBMS(DataManagement class: newDBMS)
```

Figure 5: Structure of Class Interface for Interfacing GIS and LUTS

In other words, class "Interface" can be used to cope with the following two main scenarios. One, if the same GIS shell is used, and GIS shell data and function varies with a newer version, then only the class (in the framework) that associatated with the change is rewritten. Two, if a completely new GIS shell is used then any functions associtated with new GIS shell are rewritten. Thus, the interface aims at achieving the key feature in minimising the risk of software development. In both scenarios, the structure for the class "Interface" remains the same. New instances of class "Interface" can be used for different specific implementations. For example, objects A_Interface_No1, A_Interface_No2 and A_Interface_No3 of class "Interface" can be created for linking with ARC/INFO Version 4, ARC/INFO Version 5 and GENASYS, respectively.

CONCLUSIONS

This paper has outlined a need for the integration of geographical information systems (GIS) and land-use/transport interaction (LUTI) systems, has reviewed overseas experience, and has proposed six different developmental configurations on GIS and LUTI. Five strategies for software implementation have been outlined, and some preliminary ideas on a case study have been presented. Object-oriented technology that provides an object-oriented translator between GIS and land-use/transport system (LUTS) is the recommended direction for further testing and evaluation.

With particular reference to land-use/transport planning in New South Wales, the demand for minimising data acquisition cost and sharing cost across systems suggest that to support the integration of GIS and LUTI more emphasis is required on the gaining and retaining support of high level management. The transportation strategic plan could be improved with enough data support for studying a complete system of a multimodal and coordinated transport network which includes road, rail, bus, and ferry. Ultimately, the cost for data acquisition, the ownership of data, and the management of data cannot only be supported by a single organisation. In the interim period, collaborative research between the Roads and Traffic Authority NSW, and the University of New South Wales is attempting to fill this vacuum. The following tasks are priority areas for further work: to assess transport databases and modelling; to identify suitable transport models for analysis and impact that have already been developed plus any suitable software to establish feasibility of integration with proposed systems; and to develop a prototype that demonstrates project feasibility based on a conceptual design to be formulated after the ideas in this paper have been discussed and debated.

ACKNOWLEDGEMENTS

This project is funded through the Australian Research Council (ARC) Collaborative Research Scheme. It involves the University of New South Wales - the School of Surveying (Professor John Trinder and Dr. Ewan Masters), and the School of Civil Engineering, Department of Transport Engineering - and the Roads and Traffic Authority, New South Wales (Manager Mr Brian Morrison). The views are those of the authors, and do not necessarily reflect the collective views of the research team. The authors thank Mr Roshan Aryal for his assitance with the literature survey.

REFERENCES

Bates, C.A. and Schwetz, T. (1990). "Linking Transport Planning and Geographic Information System", proc. of the 1990 Annual Conference of the Urban and Regional Information System Association I, Washington D.C., 174-185.

Bouille, F. (1992). "Decisive Criteria for GIS Comparative Study and Evaluation", Computers, Environment and Urban Systems, Vol.16, 179-185.

Black, J.A. and Ton, T.T. (1989). "Developments in Computer Aided Teaching of Land-use and Transport Planning Techniques", proc. of First International Conference on Computers in Urban Planning and Urban Management, Hong Kong, August, 51-64.

Black, J.A., Vandebona, U. and Ngo, C. (1993). Survey Report on "The Survey of the User's Experience with Transport Planning Packages", Unpublished Working Paper, Department of Transport Engineering, School of Civil Engineering, The University of New South Wales.

Cowen, D.J. (1990). "GIS versus CAD versus DBMS: What are the differences?", in D.J. Peuquet and D.F. Marble (eds), Introductory Readings in Geographic Information Systems, (London:Taylor & Francis), 52-62.

Dangermond, J. (1990) "A Classification of Software Components Commonly used in Geographic Information Systems", in D.J. Peuquet and D.F. Marble (eds), Introductory Readings in Geographic Information Systems, (London:Taylor & Francis), 30-51.

Davison, J. (1992). "JDTAP John Davison's Travel Aggregation Program, Centre of Technical Expertise in Transport Systems Modelling", VICROADS, August.

Kilsby, D., Prince, P. and Stewart, D. (1992). "Is Sketch Modelling of any Practical Value in Transport Development", proc. Australasian Transport Research Forum, Vol.17, Part 2, 197-214.

König, R. and Langbein, R. (1993). "Simulation of Traffic Management Strategies", paper for 26th International Symposium on Automobile Technology and Automation, Aachen, Germany, 13-17 September.

Lewis, S. (1990). "Use of Geographical Information Systems in Transport Modelling", ITE Journal, March, 34-37.

Nakamura, N. and Shimizu, E. (1990). "Development and Utilization of Geographical Information Systems in Urban Management - Reviewed from Examples in Japan", GIS, March, 3-9.

Nyerges, T. (1990). "Locational Referencing and Highway Segmentation in a Geographical Information System", ITE Journal, March, 27-31.

O'Neill, W. (1991). "Developing Optimal Transportation Analysis Zones Using GIS", ITE Journal, December, 33-36.

Patterson, P.E. and Ferguson, E.T. (1990). "A Dynamic Disaggregate Model of Location and Travel Behavior Using GIS", proc. of the 1990 Annual Conference of the Urban and Regional Information System Association I, Washington D.C., 186-198.

Rhind, D., Openshaw, S. and Green, N. (1988). "The Analysis of Geographical Data: Data rich, Technology Adequate, Theory Poor", in M. Rafanelli, J.C. Klensin and P. Svensson (eds), proc. 4th International Working Conference on Statistical and Scientific Database Management, Rome, 427-454.

Roberts, S.A., Gahegan, M.N., Hogg, J. and Hoyle, B. (1991). "Application of Object-oriented Databases to Geographical Information Systems", Information and Software Technology, Vol.33, No.1, January, 38-46.

Roy, G.G. and Snickars, F. (1992). "Computer-aided Regional Planning: Applications for the Perth and Helsinki Regions", Regional Science Journal, Vol.26, No.1, 79-96.

Rugg, R.D. (1992). "A Feature-Based Planning Support System", Computers, Environment and Urban Systems, Vol.16, 219-226.

Smith, J.L. (1991). "Geographic Object-Oriented Databases", Technical Report TR-HB-91-04, Division of Information Technology, CSIRO, June.

The Urban Transport Monitor (1992a). "This Week's Survey Results - Microcomputer Urban Transport Planning Software Part I", The Urban Transport Monitor, Vol.6, No.20, 9-11.

The Urban Transport Monitor (1992b). "This Week's Survey Results - Microcomputer Urban Transport Planning Software Part II", The Urban Transport Monitor, Vol.6, No.21, 9-11.

Ton, T.T. and Black, J.A. (1993). "An Object-Oriented Approach for Implementing an Integrated Four-Step Transport Planning Model", proc. of Third International Conference on Computers in Urban Planning and Urban Management, Atlanta, Georgia, July (in press).

Turner, A. (1991) "The Role of Information of Technology in Planning: Case Studies of GIS and EIS Appplications in Road Transport", in N.G. Lonergan and P.L. Simons (eds) Transport and Urban Form for Australian Cities, National Conference and Proceedings, Australian Institute of Urban Studies, N.S.W. Division, Sydney, October, 149-169.

Vonderohe, A. (1992a). "Part One - Implementing GIS at State DOTs", Public Works, February, 51-56.

Vonderohe, A. (1992b). "Part Two - Implementing GIS at State DOTs", Public Works, March, 38-42.

Wegener, M. and Junius, H. (1991). "'Universal' GIS Versus National Land Information Traditions: Internationalization of Software or Endogenous Developments?", Computers, Environment and Urban Systems, Vol.15, 219-227.

Towards a Multimedia Distributed Spatial Information System

Qiming Zhou and Allan H. Evans

School of Geography, The University of New South Wales
PO Box 1, Kensington, NSW 2033, Australia

ABSTRACT

A user-oriented, multimedia GIS is being developed in accelerating the diffusion of the technology downwards to lower level users in a range of everyday problem solving endeavours. Two critical technical problems are examined, namely, the stand-alone data base system and the developer-oriented user interface. Towards the solution of these problems, the system design of a prototype is reported to demonstrate a network-based, distributed spatial information system with multimedia user front-end. This system will provide a significant potential to make GIS technology more widely used in a great variety of disciplines with less impact on the demand of technological training that has been a prohibitive restriction for today's GIS.

INTRODUCTION

Geographical information systems (GIS) technology is now well established in an ever widening user community and its diffusion is predicted to continue apace during the next decade. Following earlier developments in the areas of environmental applications and natural resources management, the addition of new functionality is now making it possible for GIS to spread rapidly into a wide range of other application areas. This expansion is being accelerated by the development of a new generation of PC-based special purpose software with enhanced mathematical as well as cartographic modelling capabilities tailored increasingly to niche application markets including health services planning, transport, marketing, banking and finance, real estate, etc. (Maguire, *et al.*, 1991).

Thus far the diffusion of GIS has essentially been an horizontal process. Adoption of the technology is still essentially confined to professionals for higher order applications that require a solid understanding of the complexities of software packages themselves and knowledge of reasonably advanced hardware platforms. In future, however, it is expected that there will be an important vertical diffusion of the technology. It is envisaged that GIS will filter downwards to become an important tool for non-professionals to use in helping solve the kinds of problems that individuals are confronted with on an everyday basis.

This vertical diffusion of the technology will result from the efforts of current research and development across a wide front. One of the most significant areas of this research effort in making GIS more widely available to low-level users is that being undertaken into the development of multimedia systems in which images, voice, and text are fully integrated with a computing environment (Shepherd, 1991; Woelk and Kim, 1987). Technical issues, however, exist to incorporate this multimedia capability into an operational spatial information systems.

This paper reports a current research project focusing on developing an integrated multimedia spatial information system following the research in developing an integrated spatial data structure (Zhou, 1989). The significant potential that a user-oriented multimedia GIS is illustrated in accelerating the diffusion of the technology downwards to lower level users in a range of everyday problem solving endeavours. Two critical technical problems are examined, namely, the stand-alone data base system and the developer-oriented user interface. Towards the solution of these problems, the system design of a prototype is reported to demonstrate a network-based, distributed spatial information system with multimedia user front-end. The system represents a new-generation spatial data handling system, and it also yields great opportunities for potential applications which have been suffering from great difficulties in adopting spatial information technology due to data integration or/and educational issues. Together with the common strategy and standard procedure of establishing multimedia spatial data bases, the multimedia system will provide a significant potential to make GIS technology more widely used in a great variety of disciplines with less impact on the demand of technological training that has been a prohibitive restriction for today's GIS.

MULTIMEDIA SPATIAL INFORMATION SYSTEMS

Historically, software designed to manipulate spatial information has been task-specific, with clear separation between programs built for different purposes. Software packages tended to fall into one of several categories such as: *geographical information systems* (GIS) that combine cartographic data with tabular ancillary databases, *image processing systems* that manipulated image data, and *video and audio signal processing systems*. As the proliferation of spatial data processing systems increased, it becomes apparent that many applications would be best served by a combination of different data sources. Although the latest commercial software has attempted to address this need by packaging both GIS and image processing software together, this is only a temporary solution because the conceptual basis for each component still originated separately. The focus of research in spatial data analysis is therefore on the incorporation of all forms of spatial data into one truly integrated spatial information handling system (Ehlers, *et al.*, 1989).

The integration of vector cartographic data, tabular ancillary information, and raster image-based data is only an evolutionary step, consolidating components that from a geographical perspective were never separate. This concept can be extrapolated beyond static data to incorporate temporal information such as digital video and audio signals. Processes in geography, whether from an economic, human or physical perspective, occur over time. The incorporation of the temporal aspect into a spatial database is therefore a revolutionary progress in recognising that the term *process* is fundamentally a dynamic concept. Research in this field is in an early stage, partially due to the limitation and popularity of the suitable computer hardware. Some progress, however, has been reported by a number of researchers and organisations (Woelk and Kim, 1987; ESRI, 1989; Anon, 1990).

The current trend in the computer industry indicates that multimedia (or hypermedia) computer platforms will soon become general tools for information technology. Low-cost multimedia computer platforms have moved closer to commercial viability, providing the capabilities of incorporating voice, data, text, handwriting, images and video to potential applications (Figure 1). The incorporation of such multimedia systems into a carefully designed spatial information system will therefore provide the opportunity to use, manipulate and present both static and dynamic information in a truly integrated manner. This will commence a revolutionary progress in GIS development with its great potential to accelerate the downward diffusion of spatial information technology to a much larger community of end-users.

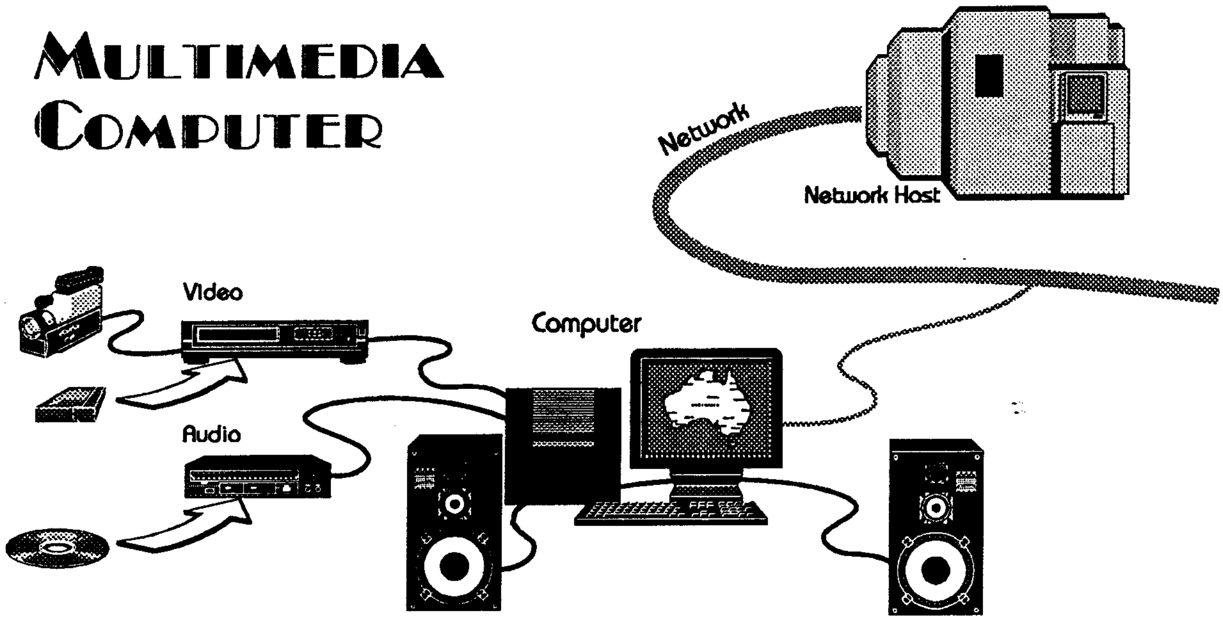


Figure 1. Multimedia computer hardware environment.

CURRENT PROBLEMS

The downward diffusion of GIS requires that user needs, rather than those of developers, must become the focus for future developments of GIS technology. Regretfully, this user orientation has not yet been given the attention it deserves. At present, the major focus of GIS development, particularly among commercial vendors is directed more towards advanced data processing algorithms and flexible tools which require a substantial commitment by users to overcome the steep learning curve in using the system. This problem is largely due to the current focus in marketing GIS to large and medium-sized organisations in which professional support in using GIS is available. This kind of support is not available to smaller groups or individuals. Therefore the development of systems suited to the needs of everyday users untrained in the technicalities of GIS will become vital.

In this regard, two major problems exist with the most current GIS software and data base systems. The first relates to spatial data accessibility; namely, stand-alone data base system design. The second relates to the ease with which GIS may be used; namely, developer-oriented user interfaces.

Stand-Alone Data Base Systems

From any user's point of viewpoint, a major benefit that a GIS provides as a decision support tool is access directly or indirectly to large volumes of spatial information at minimum cost. At present, a large proportion of the initial investment (sometimes over 80%) in establishing a GIS is that for data capture and the building of data bases. The high costs of labour, equipment and time often prevent the potential usage of GIS in many application areas, particularly those in which large volumes of up-to-date spatial data need to be retrieved but which are then used for relatively 'simple' query tasks. Adequately handling and managing large dynamic spatial data bases typically requires a high initial cost for GIS establishment and considerable hardware and software support provided by well trained personnel.

Stand-alone spatial data base design dominates current GIS software developments. Although flexible data base interfacing is often claimed by commercial vendors, the fundamental design of data base management systems (DBMS) is still predominantly focused on the manipulation of single data bases. This is particularly the case for spatial data sets because of their large volume and complexity. With this stand-alone data base design, a substantial investment in centralised facilities must be made to handle a large spatial data base (Figure 2).

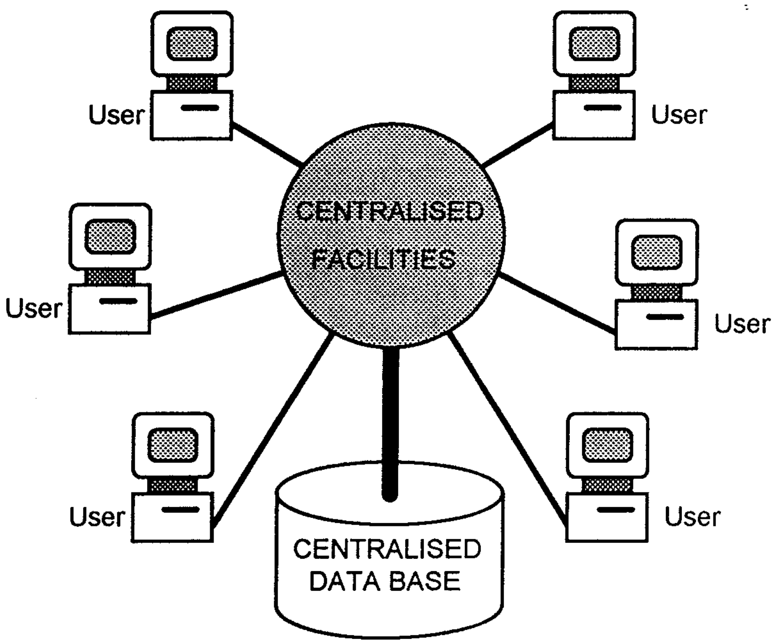


Figure 2. Centralised facilities are essential for a large stand-alone spatial data base.

If GIS technology is to be made available to a large number of low level users, this stand-alone spatial data base design is no longer sufficient. The growing emphasis on data sharing requires the establishment of large, shareable spatial data bases, the management of which is very complicated given the now well documented limitations of traditional stand-alone data base design. Accessing such a stand-alone (and in many cases, centralised) data base from many user terminals simultaneously, particularly when these are geographically dispersed over a wide area, also creates bottle-necks.

Developer-Oriented User Interfaces

Current generic GIS packages provide as many as a thousand commands or utilities to support data input, management, manipulation, analysis, and output tasks (Openshaw, 1991). They represent a "tool box" approach to providing a comprehensive capability to the user. This tool box approach, however, is designed mainly for high level users, i.e. those with an extensive technical background in GIS. It is often claimed that it is the user's (or user organisation's) responsibility to develop user-friendly front-ends for particular GIS applications; the role of GIS software vendors is essentially that of providing and packaging the basic set of tools. It is widely assumed that the 'real user' (GIS professionals) will be able to learn the details of the systems that software developers deliver. Such a developer orientation to user interfaces not only necessitates an extensive learning effort by the user to utilise a particular GIS package, but also confines user applications to that particular GIS package. In other words, the user is often 'locked in' the particular GIS package. Experience has shown that even the most experienced GIS user finds it quite difficult to 'transfer' knowledge and skills in using one package to another.

This developer-orientation to user interface design is also closely related to the current emphasis on targeting large organisations in the marketing of GIS. As part of this, it is assumed that expensive computer hardware, software and personnel will be available. Training GIS professional users is often the most costly and time consuming component of a GIS project. Because of this, the diffusion of GIS to a mass market of low level users, particularly in smaller organisations, is seriously hindered.

TOWARDS SOLUTIONS

Two approaches are proposed towards the solution of the problems outlined above. The first approach, the design and development of large, shareable, distributed data base systems, addresses the stand-alone data base problem. The second, the design and development of user-oriented graphical interfaces, attempts to overcome the problems inherent in developer-oriented user interfaces.

Large, Shareable, Distributed Data Base System

The complexities of designing a distributed computing environment has been investigated at MIT (Arfman and Roden, 1992). In the system they propose, the physical location of both the user and data is immaterial. The user's logical environment is translated into a physical environment by a special server which updates its information from a relational data base in which all system control information is recorded. Given the currently available computer technology, it is now feasible to extrapolate such a local system to a much larger one capable of operating at a national or international level. This is particularly beneficial for the developing countries because the complexity of data base management tasks can be greatly reduced and more evenly distributed and managed by personnel with only limited expertise.

The key feature of this data base system design is its distributed nature for spatial data handling. Conceptually the system consists of two components: a distributed data base and a series of local network nodes comprising hardware and software for the physical storage of spatial data. The distributed data base system has a central data manager and a relational data base containing information about data stored at each local network node. No data is necessarily stored at the central site which acts primarily as a means of locating and accessing data stored at the individual nodes. Each node in the network may consist of a single computer or a local area network (Figure 3).

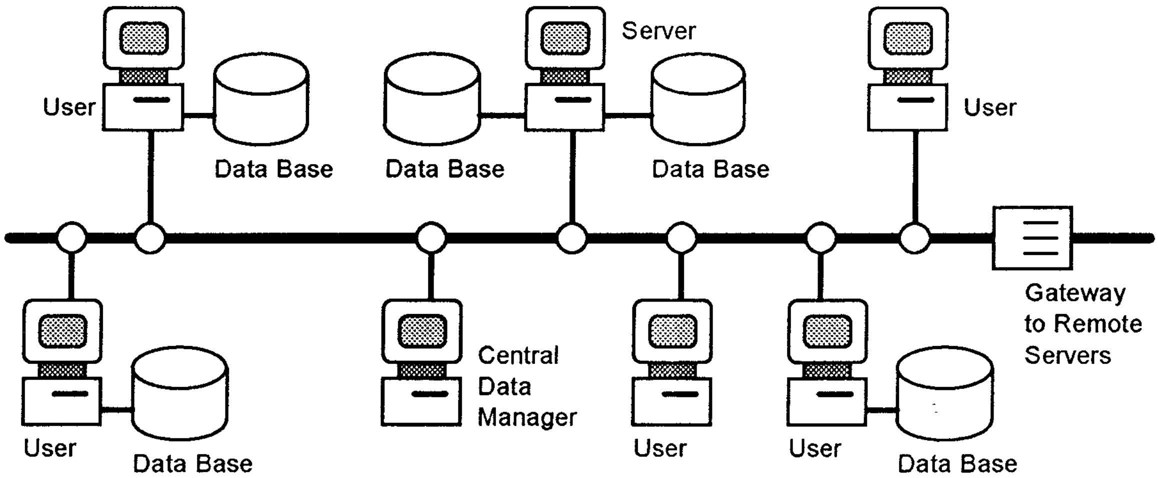


Figure 3. A distributed information system with a central data manager and networked data bases.

The benefits of distributed data base systems over their conventional stand-alone counterparts are significant:

1. There is no need to establish a massive central facility, potentially subject to high running costs, to maintain a large spatial data base. Since the complex data base management tasks are undertaken by the central data manager, end-users will be supported with well-maintained, shared spatial data thus obviating the need to establish a stand-alone data base.

2. The traffic related to large volumes of spatial data transfer are more evenly distributed over the network, since bottle-necks in accessing the centralised data base may be avoided. This will guarantee easy access to large spatial data bases by potential users over a wide area.

3. The user's accessibility to and response from a distributed information system will be greatly improved compared to a centralised data base system which is frequently overloaded as a result of the large number of user queries over a wide area.

4. They provide the potential for easy implementation of a modular processor structure to the distributed data base system. Therefore not only data base access but also data processing tasks may also be distributed over the network to make more efficient uses of the computing infrastructure.

User-Oriented Graphic Interfaces

Given the state-of-art of computer technology, it is now possible to develop better user-oriented system interfaces for GIS. Computers are now widely used on a day-today basis given the substantial and varied collection of general computing software currently available. Therefore, it is reasonable to expect that potential GIS users have some familiarity with and skills in using some kind of general computing software, such as a word processor. The challenge, therefore, is to develop powerful GIS engines with a simple-to-use interface which can be speedily mastered by the end-user.

Significant progress has been reported in developing graphic user interfaces (GUI) and spatial query languages (Frank and Mark, 1991). This progress, however, is still essentially developer-oriented with the result that interfaces are still far from being 'user-friendly' compared with to those in the range of general computing software currently available for word processing, spread sheeting, and desktop publishing.

A 'user-friendly' interface for GIS must have a strong user rather than developer orientation. In other words, vendors have to fit the software into a framework that potential users are already familiar with and already used to. From a user point of view, the interface for GIS should provide the following capabilities to facilitate its use by ordinary users (Evans and Zhou, 1992):

- a device-independent GUI with a similar 'look and feel' to that of standard GUI's which have been proposed for more general software packages;

- a comprehensive tool box for map editing and data processing giving the user freedom in modifying graphic objects interactively;

- dynamic links between spatial objects, images, attributes, and, potentially, digital video and audio information enabling graphical as well as logical query of the data entities in the data base; and

- the capability to record, edit and play back macros, and the automatic generation of spatial models which allow further development of automated spatial data processing procedures.

A number of common standards that provide the necessary design basis are already in place such as the *Common User Access* (CUA) *advanced interface for software applications* proposed by IBM's *Systems Application Architecture* (SAA) (IBM, 1989), *Openlook* by Sun Microsystems (Sun Microsystems, 1990) and *Motif* by the Open System Foundation. Based on these standards, a user-oriented GIS interface can be developed relatively easily to present a consistent front for within and across applications.

In designing a user-oriented GIS interface, a prototype has been developed to implement the SAA CUA interface (Evans and Zhou, 1992). To incorporate spatial modelling capabilities into the GIS, an English-like macro language has also been developed based on the GIS data base engine. This prototype is currently being extended further to incorporate multimedia capabilities for the presentation and visualisation of spatial data (Zhou and Evans, 1992).

The design of the prototype is structured at three conceptual levels (Figure 4):

- User Level. A consistent front-end which is independent of the operating system and data type. A common user interface is developed with modeless windows, icons, mouse-driven features and popup menus as specified by the CUA for SAA.

- Data Level. Device-independent data query and processing algorithms and the interface between different data types. This includes data capabilities for handling raster and vector data, multimedia data, and attribute information managed by a relational data base engine.

- Virtual Device Level. A device interface designed to communicate with a standard operating environment such as *Microsoft Windows* on MS-DOS based PC and *X Window* on UNIX platforms.

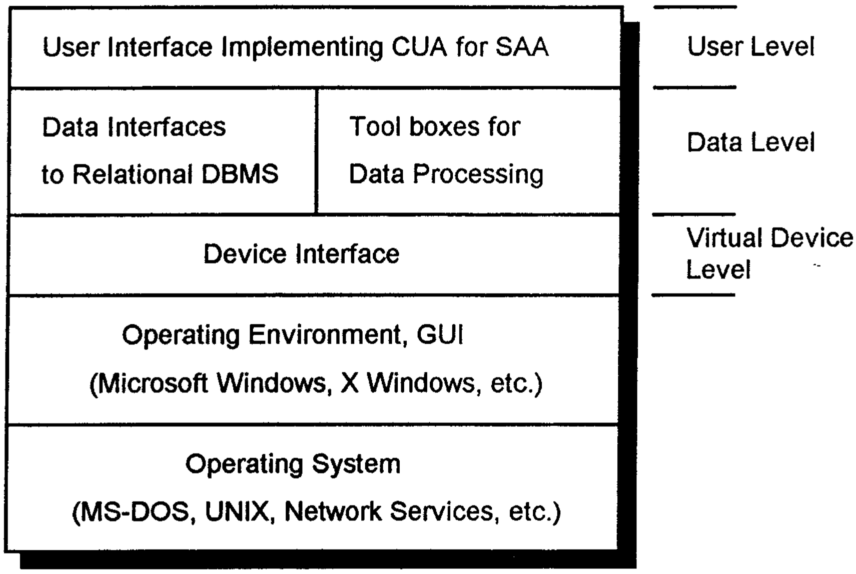


Figure 4. Conceptual levels of a user-oriented GIS interface.

Conceptually the prototype GIS is intended to be a *map processor* that operates in the same way as a word processor. With this prototype system, a map (or spatial data layer) is treated similarly to the way that a document is treated by a word processor. The user may open, close or create a map; add, delete or modify spatial objects; hear and view audio and video clips and multimedia documents, or search or query spatial and attribute data bases which may be handled by a networked DBMS.

CONCLUSION

We believe that one of the most significant developments in GIS in the coming decade will be the design of systems, particularly multimedia ones, capable of supporting everyday spatial decision-making by low level users. It is not unreasonable to predict that by the year 2000 GIS has the potential of becoming as ubiquitous in the community and as easy to use as word processing and spread sheet packages are today. Realisation of this however, will necessitate a fundamental shift in current practice away from an emphasis on the design of stand-alone data base systems and user interfaces that are essentially developer rather than user oriented. The vertical diffusion of GIS technology downwards as a support tool for decision-making by non-expert users will be predicated on the design and development of large distributed data bases which users can share, and especially the design of graphic interfaces that are easy to use. These topics are currently being addressed by ongoing research at the University of New South Wales as significant research topics in their own right in connection with the design and development of the *Asian Spatial Information Analysis Network* which will be used by low level users for Asian studies at Australian tertiary academic institutions to access spatial and other data for the countries of Asia. They are also being addressed as part of a longer term research program that focuses on the development of easy to use multimedia systems for instructional and operational purposes by university students, and as decision support tools suitable for use by ordinary people in everyday situations.

REFERENCES

Anon (1990), "The French revolution of 1989", Mapping Awareness, Vol. 4, No. 9, 48-49.

Arfman, J.M. and Roden, P. (1992), "Project Athena: supporting distributed computing at MIT", IBM Systems Journal, Vol. 31, 550-563.

Ehlers, M., Edwards, G. and Bedard, Y. (1989), "Integration of remote sensing with geographical information systems: a necessary evolution", Photogrammetric Engineering and Remote Sensing", Vol. 55, No. 11, 1619-1627.

ESRI (1989), "Integration of geographic information technologies", ARC News, Winter, 24-25.

Evans, A.H. and Zhou, Q. (1992), "Design of a graphic user interface for an integrated geographical information system", Proceedings of the 4th Australasian Remote Sensing Conference, Wellington, November, Vol. 2, 66-72.

Frank, A.U. and Mark, D.M. (1991), "Language issues for GIS", in Maguire, D., Goodchild, M., and Rhind, D. (eds), Geographical Information Systems: Principles and Applications, Longman Scientific & Technical, London, 147-163.

IBM (1989), Systems Application Architecture: Common User Access Advanced Interface Design Guide, IBM Corp.

Maguire, D., Goodchild, M., and Rhind, D. (eds) (1991), Geographical Information Systems: Principles and Applications. Longman Scientific & Technical, London.

Openshaw, S. (1991), "Developing appropriate spatial analysis methods for GIS", in Maguire, D., Goodchild, M., and Rhind, D. (eds), Geographical Information Systems: Principles and Applications, Longman Scientific & Technical, London, 389-402.

Shepherd, I.D.H. (1991), "Information integration and GIS", in Maguire, D., Goodchild, M., and Rhind, D. (eds), Geographical Information Systems: Principles and Applications, Longman Scientific & Technical, London, 337-360.

Sun Microsystems (1990), OPENLOOK Graphical User Interface Application Style Guidelines, Addison-Wesley, Reading.

Woelk, D., and Kim, W. (1987), "Multimedia information management in an object oriented database system", Proceedings, 13th Very Large Databases Conference, Brighton, UK, 319-329.

Zhou, Q. and Evans, A.H. (1992), "Development of a multimedia spatial information system", Proceedings of the 17th Congress of International Society for Photogrammetry and Remote Sensing, Washington D.C., August, Part B2, 67-71.

Zhou, Q. (1989), "A method in integrating remote sensing and geographical information systems", Photogrammetric Engineering and Remote Sensing, Vol. 55, No. 5, 591-596.

THE INTEGRATION OF GIS AND REMOTE SENSING IN REGIONAL ECONOMIC PLANNING

Qiming Zhou

School of Geography, The University of New South Wales
PO Box 1, Kensington, NSW 2033, Australia

ABSTRACT

An integrated GIS has been developed as the core of an Asian Development Bank funded project which will provide the foundation for the regional economic planning in semi-arid areas in North China. The GIS is composed of a regional geographical data base and a collection of spatial models. The regional geographical data base contains data sets from terrestrial-based sources including thematic maps and statistical records, and remote sensed data. A series of spatial models have been developed by experts in various disciplines such as soil erosion, economics, rural development and arid land management. These models are then implemented into the GIS environment and an extensive mapping tasks are conducted. As the result, a regional development plan is produced to balance the demand from different interest group of land use to achieve an optimum, environmentally sound economic development.

INTRODUCTION

Desertification and land degradation in arid and semi-arid areas has been one of the most significant environmental problems in China. More than half of China's territory has an annual precipitation below 400 mm. Within this area landforms are complex and constantly changing, and accessibility is limited. Although considerable progress has been made in monitoring changes in this difficult environment (Zhu and Wang, 1991), the Chinese government identified the need for additional technology transfer in the areas of remote sensing and Geographical Information Systems (GIS) to help the plan of long term management strategies to prevent further land degradation.

Recently, a Asian Development Bank (ADB) funded pilot research project has commenced to develop technology for regional economic development planning in a dryland area in North China, where the ecosystem is extremely unstable and sensitive to human land use activities. Due to the fact that a huge storage of mineral resources have recently been discovered which may turn this area into one of the largest coal mining area in the World, the regional economic planner is facing a complicated task to balance the demand on lands by traditional agriculture and newly developed mining industry, and the growth of population in this area. To provide a useful resolution, the pilot project has taken a comprehensive approach to incorporate multi-disciplinary expertise of local and international specialists in the areas of arid land management, economics, rural planning, soil sciences, remote sensing and GIS. The goal of the project is to produce a regional economic development plan which will be used as the guideline for optimum land use in the fragile ecosystem.

This paper reports the technical design of the project focusing on the operational approach by which GIS and remotely sensed data are integrated for comprehensive economic planning and land resource management. The proposed methodology to achieve this integration is task-oriented emphasising the optimum cost/benefit ratio which is particularly important for projects in developing countries.

STUDY AREA

The study area for this project is located in North China at the border of Shanxi and Shaanxi Provinces and Inner Mongolia (approximately 600 km west of Beijing). The project covers an area about 50,000 km^2 in eight counties with a population of around four million people.

The study area is located in the interchange zone between arid sandy desert/grassland (Mao-wu-su sand field) and semi-arid loess plateau (Li, 1990). The elevation of the most part of the area is between 1,000 - 1,500 m ASL with annual rainfall between 250 - 450 mm. The north part of the region is the Mao-wu-su sand field where grasslands and aeolian landforms such as sand dunes are predominant. Toward south is the loess plateau where the landforms are predominantly controlled by the surface material - loess - and extremely unevenly distributed rainfall. Yellow river cuts through the northern and eastern sides of the region creating a local relief of 400 - 500 m (Figure 1).

The traditional land use activities are animal husbandry in the northwest and agriculture in the southeast. Recently huge mineral (coal) reserves have been discovered in this region and large investments have been poured in by central government and foreign investors, which gradually build this area as one of the largest coal mining area in the world.

Environmental problems are fundamental for the regional development planning. The region has long faced the critical problem to balance the growing population and limited land resources. Due to the long history of uncontrolled agricultural land use, the desertification and land degradation has been a very serious problem leading to the permanent damage of the fragile ecosystem. Among numerous land management issues, the most critical problems include:

- desertification in grassland area resulting in the invasion of mobile sand dunes and the deduction of vegetative cover;
- pasture degradation resulting in decreasing high-quality grazing species and increasing harmful species;
- salinity in irrigated lands particularly in the area along Yellow River;
- severe soil erosion and mass movement in loess area;
- industrial pollution and environmental problems such as accelerated mass movement and river sedimentation; and
- rapid urbanisation resulting in loss of the best agricultural lands and increased demand on water resources and water pollution.

In addition to these problems, drought has been one of the most frequent threats to the regional economic development. As rainfall is extremely unstable in this area, large proportion of annual rainfall falls during the summer time (June - August), usually with a very high rainfall

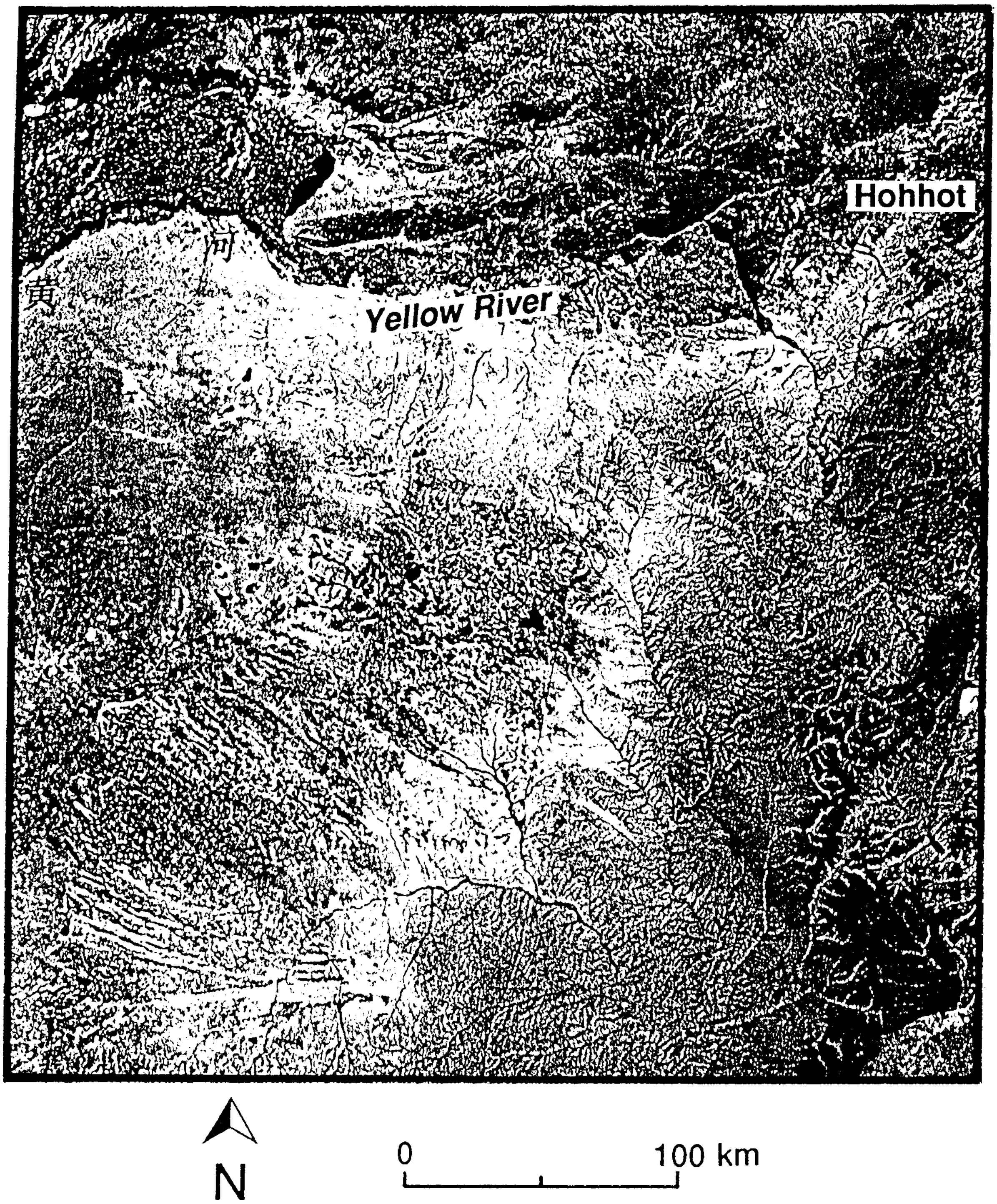


Figure 1. The study area is located in the interchange zone between arid sand desert/grassland and semi-arid loess plateau (from "Landsat Image Map of China", Chinese Academy of Sciences, 1991).

intensity. This results in high temporary run-off, washing away the fertile top soils into Yellow River system.

Facing critical problems as shown above, the regional planners have to seek the answers to many vital questions in their decision-making such as:

- What is the current vegetation productivity which will determine the stock rate for pastoral lands?
- What is the optimum balance between industrialisation, which will bring the wealth to the region, and environmental protection, which will prevent the irreversible damage to the environment?
- Where is the area for urban growth without losing valuable agricultural lands?
- What is the best approach to use the massively over-supplied labour force due to the rapid population growth without causing more environmental damage?

METHODOLOGY

In order to answer above questions, this project has selected an comprehensive technical approach to incorporate technologies from traditional studies, such as soil erosions, land use and land capability mapping and rural planning, and those from new technologies, such as GIS and remote sensing.

An integrated GIS has been developed as the core of the project which will provide the foundation for the regional economic planning. The GIS is composed of a regional geographical data base and a collection of spatial models. The regional geographical data base contains a number of thematic map layers including land use, land systems, administrative boundaries, infrastructures and drainage; remote sensed data including Landsat TM, SPOT and NOAA AVHRR images; digital terrain models; and a comprehensive collection of statistical records. The data base covers an area of approximate 50,000 km^2 with data retrieved from 43 1:100,000 map sheets. A series of spatial models have been developed by experts in various disciplines such as soil erosion, economics, rural development and arid land management. These models are then implemented into the GIS environment and an extensive mapping tasks are conducted. As the result, a regional development plan is produced to balance the demand from different interest group of land use to achieve an optimum, environmentally sound economic development.

A 'task-oriented' approach (Ehlers, *et al.*, 1989) has been taken towards the solution of the management problems. The regional information system has been constructed in two levels, namely expert level and user level. The system at the expert level has been established in the Institute of Remote Sensing Applications (IRSA) at the Chinese Academy of Sciences (CAS), where a selection of specialists and technician work together to carry on complicated tasks and package the technology of the integration of remote sensing and GIS. At the user level, PC-based low cost systems are installed in the regional office with the integrated data and spatial model package to allow daily operations to support decision-making in land resource management.

A heterogeneous hardware and software environment has been established at the expert level in IRSA, in which both PC and UNIX workstation hardware has been networked with shareable peripherals. PC-based ILWIS (ITC, 1992) and workstation-based ARC/INFO software packages have been used to carry out tasks of data input and manipulation.

Four stages of the research and development project have been identified including data input, data integration, spatial and economic modelling and local implementation.

Data Input

Six layers of terrestrial-based spatial data have been digitised using the PC-based ILWIS system including land use, land systems, river network, road network, residential and industrial sites and administrative boundaries. In addition to these, digital elevation data have also been acquired from the National Laboratory of Environment and Resource Information Systems (LREIS) at CAS. Since the accessibility to workstation based GIS was limited, the data input tasks are mainly carried out in PC-based environment. The data sets, however, are merged later on workstations because of the heave demand on data processing power (Figure 2).

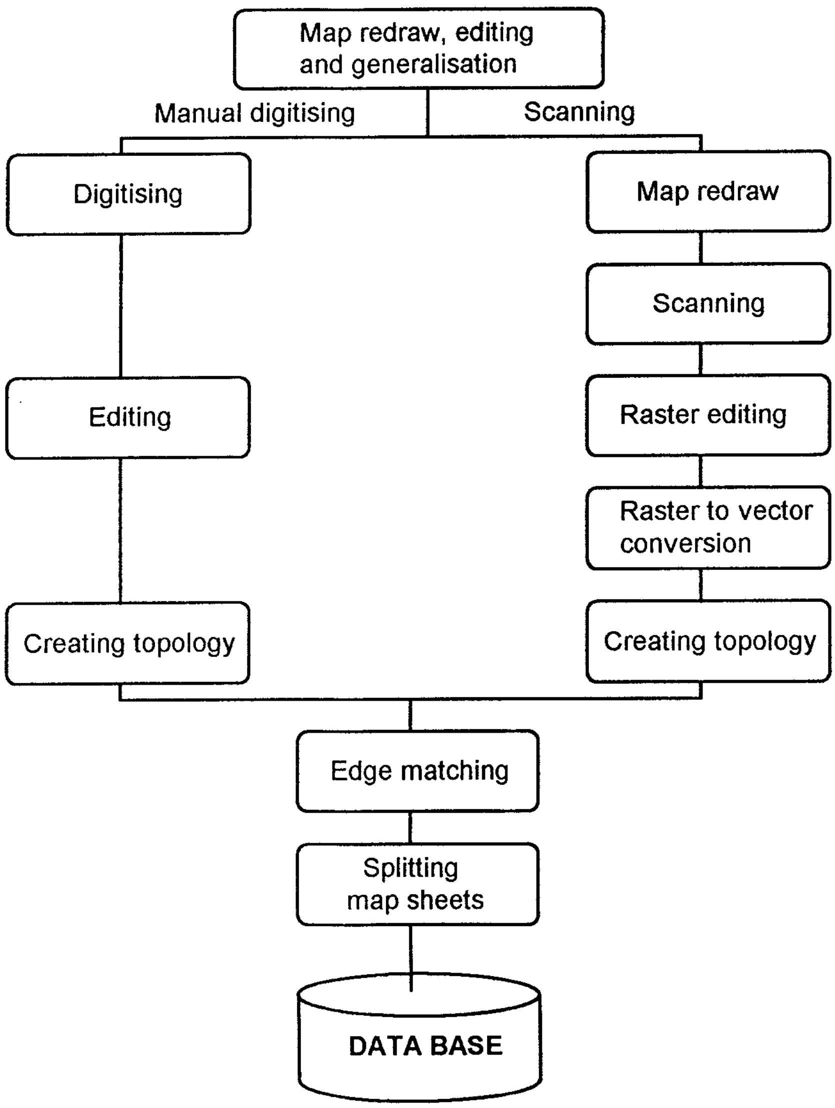


Figure 2. Data input procedure for terrestrial-based thematic maps.

Digital remote sensing images have been acquired including Landsat TM images covering the entire study area, SPOT images covering selected pilot study sites and a number of NOAA AVHRR scenes for resource monitoring. These images have been geometrically rectified and registered on the 1:100,000 map sheets.

Statistical records have been acquired from local and central governments giving critical information for regional economic planning. The data are stored as relational tables and pointers have been created to link the statistical information to its corresponding spatial location.

Data Integration

All spatial data including digitised thematic maps and remote sensing images and tabular data have been integrated into a single information system and spatial indices have been created to link different data sets together using there spatial locations (Figure 3). Based on this, an integrated spatial data base has initially been established using workstation-based ARC/INFO system and all digitised map sheets have been merged together to create an entire scene for the region. Since the data sets turn into large digital data volume, manipulation of these data is beyond the capacity of a PC-based GIS.

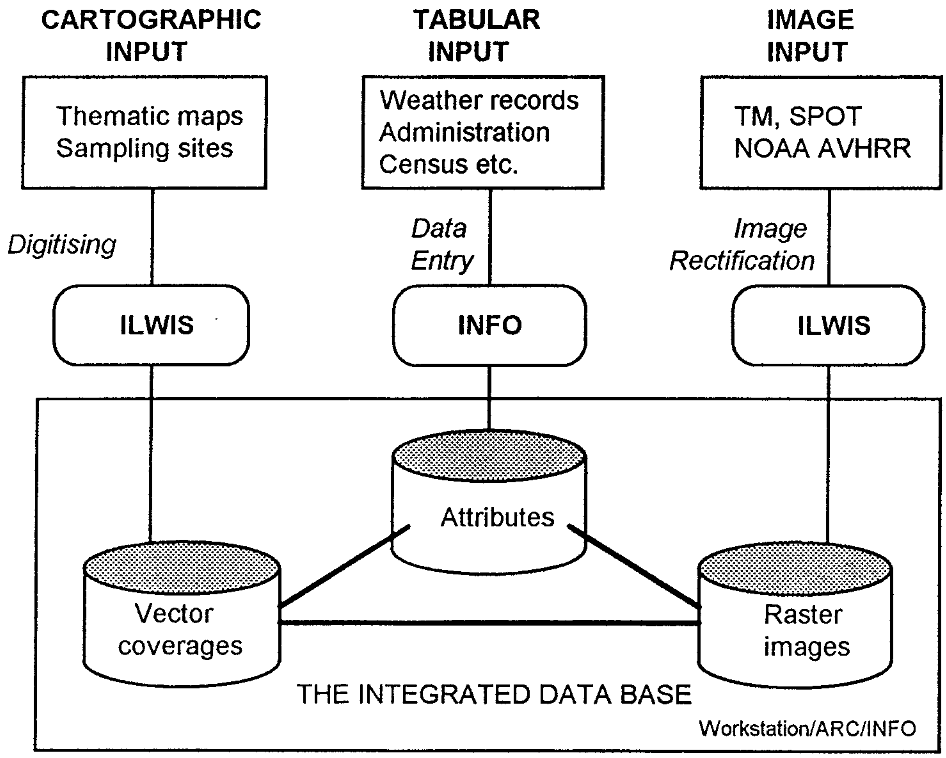


Figure 3. The structure of the integrated spatial information system.

Spatial and Economic Modelling

The integrated spatial information system provides a powerful set of tools to specialists in rural management, land use, geomorphology, soil and regional economic planning for their tasks in assessing the major causes of desertification and land degradation, in evaluating current land conservation activities, and in producing environmentally sound regional economic development plan.

Spatial and economic models have been developed by application specialists to answer specific questions regarding environment, resource and human activities. The models are then interpreted and simplified into computational procedures that are based on data processing tools provided by GIS and spatial and attribute data stored in the integrated data base. The processing results are then re-evaluated by the application specialists for accuracy and usability assessment.

Multitemporal remotely sensed images are particularly important in this project. Relative simple image processing procedures have been developed to convert the raw images into more usable forms for land resource management. Based on the initial detailed ground truth studies, the remotely sensed data provide a reliable data source for updating, particularly for the pastoral area located in the northwest of the region.

Local Implementation

It is obvious that awareness of the benefit brought from GIS and remote sensing technology plays a key role for the technological implementation to the local area. Therefore, all data processing tasks have been carried out with the close liaison to decision makers in local and central governments (Figure 4). A series of training programs have been conducted to introduce the technology to the local operators and government officials, and local supporting centres are established in order to provide adequate technical services to the local decision-makers.

In implementing the technology in local management, the spatial data base and modelling procedures are packaged into low-cost PC-based operating environment. The large data volumes have been sub-divided into sub-areas based on 1:100,000 map sheets and supplied to the regional land management authorities. The data processing procedures are also automated and packaged into 'turn-key' tasks so that the local operators with short-term training can adequately undertake daily land resource management tasks.

CONCLUDING REMARKS

This project has demonstrated a practical use of GIS and remote sensing technology in a developing nation. It has also exposed the great potential and usefulness of the integration of GIS and remote sensing technology in regional economic development plan, particularly in areas where the ecosystem is fragile, land productivity is relatively low, and conflicts in land use have occurred.

One of the most significant challenge to the GIS and remote sensing technologies is the environment monitoring and change prediction at the practical operational costs. This is particularly important in developing countries, such as China, where limited resources are available. To achieve a sound cost/benefit ratio, the technological development must focus on the provision of operational benefit to the local management, where the adequate technical support should not be expected. It is therefore extremely important for a successful implementation to package the technology to assist the low-level users in their daily management tasks.

Further research should be focused on more advanced methods in data integration, better packaging of the technology, more comprehensive spatial and economic models, and methods to adopt low-cost remotely sensed data in real-time land monitoring.

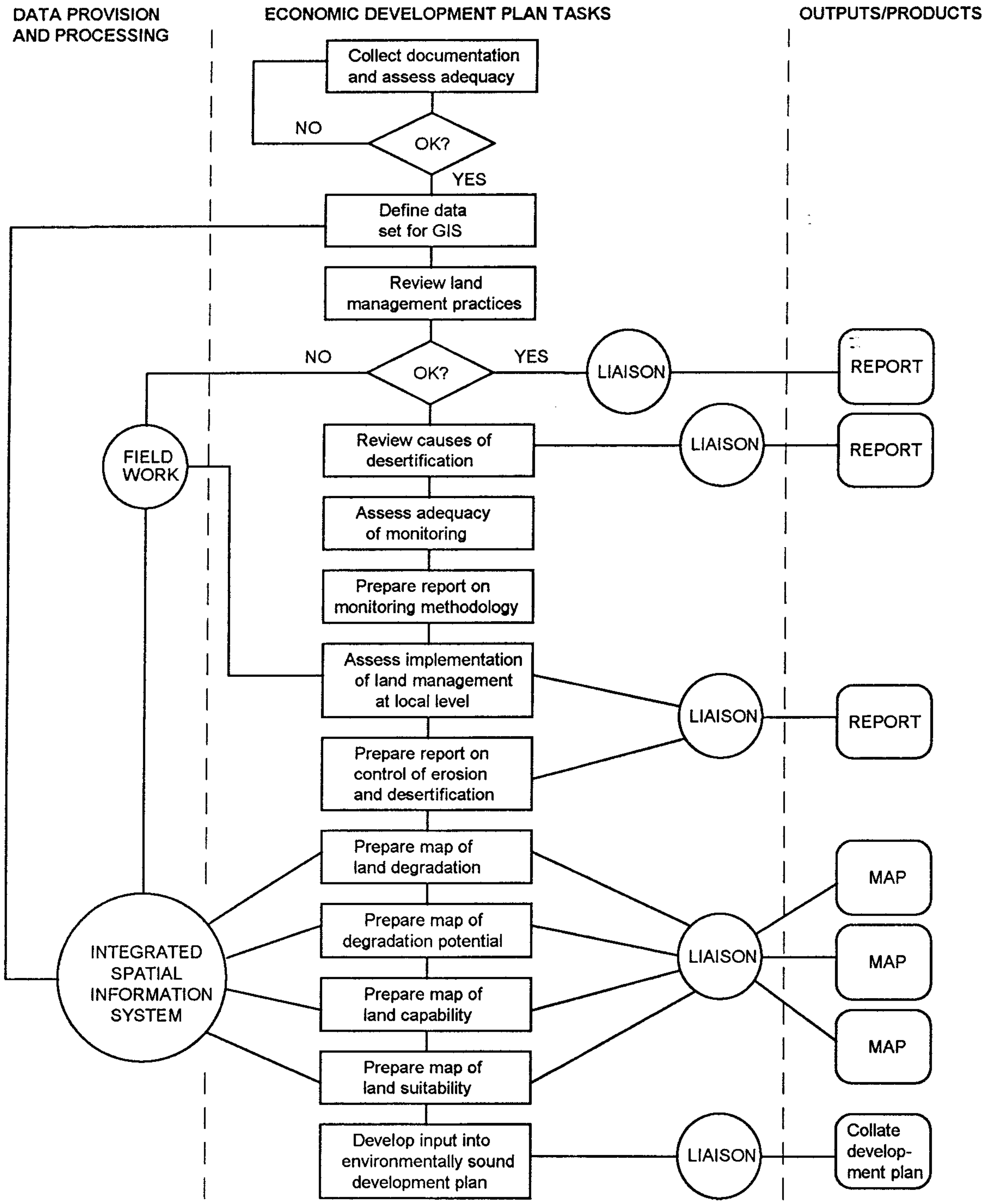


Figure 4. Implementation of the integrated spatial information system in regional economic development plan (after Squires, 1992).

REFERENCES

Chinese Acedemy of Sciences (1991), Landsat Image Map of China, Science Press, Beijing.

Publications from

THE SCHOOL OF SURVEYING, THE UNIVERSITY OF NEW SOUTH WALES.

All prices include postage by surface mail. Air mail rates on application. (Effective Oct. 1992)

To order, write to Publications Officer, School of Surveying, The University of New South Wales, P.O. Box 1, Kensington N.S.W., 2033 AUSTRALIA

NOTE: ALL ORDERS MUST BE PREPAID

UNISURV REPORTS - G SERIES

Price (including postage): $6.00

G14. A. Stolz, "The computation of three dimensional Cartesian coordinates of terrestrial networks by the use of local astronomic vector systems", Unisurv Rep. 18, 47 pp, 1970.

G16. R.S. Mather et al, "Communications from Australia to Section V, International Association of Geodesy, XV General Assembly, International Union of Geodesy and Geophysics, Moscow 1971", Unisurv Rep. 22, 72 pp, 1971.

G17. Papers by R.S. Mather, H.L. Mitchell & A. Stolz on the following topics:- Four-dimensional geodesy, Network adjustment and Sea surface topography, Unisurv G17, 73 pp, 1972.

G18. Papers by L. Berlin, G.J.F. Holden, P.V. Angus-Leppan, H.L. Mitchell & A.H. Campbell on the following topics:- Photogrammetry co-ordinate systems for surveying integration, Geopotential networks and Linear measurement, Unisurv G18, 80 pp, 1972.

G19. R.S. Mather, P.V. Angus-Leppan, A. Stolz & I. Lloyd, "Aspects of four-dimensional geodesy", Unisurv G19, 100 pp, 1973.

G20. Papers by J.S. Allman, R.C. Lister, J.C. Trinder & R.S. Mather on the following topics:- Network adjustments, Photogrammetry, and 4-Dimensional geodesy, Unisurv G20, 133 pp, 1974.

G21. Papers by E. Grafarend, R.S. Mather & P.V. Angus-Leppan on the following topics:- Mathematical geodesy, Coastal geodesy and Refraction, Unisurv G21, 100 pp, 1974.

G22. Papers by R.S. Mather, J.R. Gilliland, F.K. Brunner, J.C. Trinder, K. Bretreger & G. Halsey on the following topics:- Gravity, Levelling, Refraction, ERTS imagery, Tidal effects on satellite orbits and Photogrammetry, Unisurv G22, 96 pp, 1975.

G23. Papers by R.S. Mather, E.G. Anderson, C. Rizos, K. Bretreger, K. Leppert, B.V. Hamon & P.V. Angus-Leppan on the following topics:- Earth tides, Sea surface topography, Atmospheric effects in physical geodesy, Mean sea level and Systematic errors in levelling, Unisurv G23, 96 pp, 1975.

G24. Papers by R.C. Patterson, R.S. Mather, R. Coleman, O.L. Colombo, J.C. Trinder, S.U. Nasca, T.L. Duyet & K. Bretreger on the following topics:- Adjustment theory, Sea surface topography determinations, Applications of LANDSAT imagery, Ocean loading of Earth tides, Physical geodesy, Photogrammetry and Oceanographic applications of satellites, Unisurv G24, 151 pp, 1976.

G25. Papers by S.M. Nakiboglu, B. Ducarme, P. Melchior, R.S. Mather, B.C. Barlow, C. Rizos, B. Hirsch, K. Bretreger, F.K. Brunner & P.V. Angus-Leppan on the following topics:- Hydrostatic equilibrium figures of the Earth, Earth tides, Gravity anomaly data banks for Australia, Recovery of tidal signals from satellite altimetry, Meteorological parameters for modelling terrestrial refraction and Crustal motion studies in Australia, Unisurv G25, 124 pp, 1976.

G26. Papers by R.S. Mather, E.G. Masters, R. Coleman, C. Rizos, B. Hirsch, C.S. Fraser, F.K. Brunner, P.V. Angus-Leppan, A.J. McCarthy & C. Wardrop on the following topics:- Four-dimensional geodesy, GEOS-3 altimetry data analysis, analysis of meteorological measurements for microwave EDM and Meteorological data logging system for geodetic refraction research, Unisurv G26, 113 pp, 1977.

G27. Papers by F.K. Brunner, C.S. Fraser, S.U. Nasca, J.C. Trinder, L. Berlin, R.S. Mather, O.L. Colombo & P.V. Angus-Leppan on the following topics:- Micrometeorology in geodetic refraction, LANDSAT imagery in topographic mapping, adjustment of large systems, GEOS-3 data analysis, Kernel functions and EDM reductions over sea, Unisurv G27, 101 pp, 1977.

G29. Papers by F.L. Clarke, R.S. Mather, D.R. Larden & J.R. Gilliland on the following topics:- Three dimensional network adjustment incorporating ξ , η and N, Geoid determinations with satellite altimetry, Geodynamic information from secular gravity changes and Height and free-air anomaly correlation, Unisurv G29, 87 pp, 1978.

From June 1979 Unisurv G's name was changed to Australian Journal of Geodesy, Photogrammetry and Surveying. These can be ordered from The Managing Editor, Australian Journal of Geodesy, Photogrammetry and Surveying, Institution of Surveyors - Australia , Nos 27 - 29 Napier Close, Deakin, ACT 2600, AUSTRALIA.

UNISURV REPORTS - S SERIES

S8 - S20	Price (including postage):		$10.00
S27 onwards	Price (including postage):	Individuals	$25.00
		Institutions	$30.00

S8 A. Stolz, "Three-D Cartesian co-ordinates of part of the Australian geodetic network by the use of local astronomic vector systems", Unisurv Rep. S 8, 182 pp, 1972.

S10 A.J. Robinson, "Study of zero error & ground swing of the model MRA101 tellurometer", Unisurv Rep. S 10, 200 pp, 1973.

S12. G.J.F. Holden, "An evaluation of orthophotography in an integrated mapping system", Unisurv Rep. S 12, 232 pp, 1974.

S14. Edward G. Anderson, "The Effect of Topography on Solutions of Stokes` Problem", Unisurv Rep. S 14, 252 pp, 1976.

S16. K. Bretreger, "Earth Tide Effects on Geodetic Observations", Unisurv S 16, 173 pp, 1978.

S17. C. Rizos, "The role of the gravity field in sea surface topography studies", Unisurv S 17, 299 pp, 1980.

S18. B.C. Forster, "Some measures of urban residential quality from LANDSAT multi-spectral data", Unisurv S 18, 223 pp, 1981.

S19. Richard Coleman, "A Geodetic Basis for recovering Ocean Dynamic Information from Satellite Altimetry", Unisurv S 19,332 pp, 1981.

S20. Douglas R. Larden, "Monitoring the Earth's Rotation by Lunar Laser Ranging", Unisurv Report S 20, 280 pp, 1982.

S27. Bruce R. Harvey, "The Combination of VLBI and Ground Data for Geodesy and Geophysics", Unisurv Report S27, 239 pp, 1985.

S29. Gary S. Chisholm, "Integration of GPS into hydrographic survey operations", Unisurv S29, 190 pp, 1987.

S30. Gary Alan Jeffress, "An investigation of Doppler satellite positioning multi-station software", Unisurv S30, 118 pp, 1987.

S31. Jahja Soetandi, "A model for a cadastral land information system for Indonesia", Unisurv S31, 168 pp, 1988.

S32. D. B. Grant, "Combination of terrestrial and GPS data for earth deformation studies" Unisurv S32, 285 pp, 1990.

S33. R. D. Holloway, "The integration of GPS heights into the Australian Height Datum", Unisurv S33, 151 pp.,1988.

S34. Robin C. Mullin, "Data update in a Land Information Network", Unisurv S34, 168 pp. 1988.

S35. Bertrand Merminod, "The use of Kalman filters in GPS Navigation", Unisurv S35, 203 pp., 1989.

S36. Andrew R. Marshall, "Network design and optimisation in close range Photogrammetry", Unisurv S36, 249 pp., 1989.

S37. Wattana Jaroondhampinij, "A model of Computerised parcel-based Land Information System for the Department of Lands, Thailand," Unisurv S37, 281 pp., 1989.

S38. C. Rizos (Ed.), D.B. Grant, A. Stolz, B. Merminod, C.C. Mazur "Contributions to GPS Studies", Unisurv S38, 204 pp., 1990.

S39. C. Bosloper, "Multipath and GPS short periodic components of the time variation of the differential dispersive delay", Unisurv S39, 214 pp., 1990.

S40. John Michael Nolan, "Development of a Navigational System utilizing the Global Positioning System in a real time, differential mode", Unisurv S40, 163 pp., 1990.

S41. Roderick T. Macleod, "The resolution of Mean Sea Level anomalies along the NSW coastline using the Global Positioning System", 278 pp., 1990.

S42. Douglas A. Kinlyside, "Densification Surveys in New South Wales - coping with distortions", 209 pp., 1992.

S43. A. H. W. Kearsley (ed.), Z. Ahmad, B. R. Harvey and Adolfientje Kasenda, "Contributions to Geoid Evaluations and GPS Heighting", 209 pp. + xii, 1993.

PROCEEDINGS

Prices include postage by surface mail

P1. P.V. Angus-Leppan (Editor), "Proceedings of conference on refraction effects in geodesy & electronic distance measurement", 264 pp., 1968. Price: $10.00

P2. R.S. Mather & P.V. Angus-Leppan (Eds), "Australian Academy of Science/International Association of Geodesy Symposium on Earth's Gravitational Field & Secular Variations in Position", 740 pp., 1973. Price $15.00

SPECIAL GPSCO PUBLICATION

Price includes postage by surface mail

GPSCO 1. Simon McElroy, Ewan Masters, Glenn Jones, Douglas Kinlyside, Chris Rizos, Adrian Siversten, Patrick Brown, Owen Moss, Greg Dickson, "Getting Started with GPS Surveying", 186 pp, Approx 90 diagrams, 1992. Price $30.00

MONOGRAPHS

Prices include postage by surface mail

M1.	R.S. Mather, "The theory and geodetic use of some common projections", (2nd edition), 125 pp., 1978.	Price	$15.00
M2.	R.S. Mather, "The analysis of the earth's gravity field", 172 pp., 1971.	Price	$8.00
M3.	G.G. Bennett, "Tables for prediction of daylight stars", 24 pp.,1974.	Price	$5.00
M4.	G.G. Bennett, J.G. Freislich & M. Maughan, "Star prediction tables for the fixing of position", 200 pp., 1974.	Price	$8.00
M8.	A.H.W. Kearsley, "Geodetic Surveying", 77 pp., 1988.	Price	$12.00
M.9	R.W. King, et al, "Surveying with GPS", 133 pp., 1985.	Price	$20.00
M10.	W. Faig, "Aerial Triangulation and Digital Mapping", 102 pp., 1986.	Price	$16.00
M11.	W.F. Caspary, "Concepts of Network and Deformation Analysis", 183 pp., 1988.	Price	$25.00
M12.	F.K. Brunner, "Atmospheric Effects on Geodetic Space Measurements", 110 pp., 1988.	Price	$16.00
M13.	Bruce R. Harvey, "Practical Least Squares and Statistics for Surveyors", 229 pp., 1990.	Price	$25.00
M14.	Ewan G. Masters & John R. Pollard (Ed.), "Land Information Management", 269 pp., 1991. (Proceedings LIM Conference, July 1991).	Price	$50.00