Ordering
All product requests should be directed to:

ORDER DESK
Prince Albert Satellite Station
Energy, Mines and Resources Canada
P.O. Box 1150
Prince Albert, Saskatchewan
S6V 5S7
Telephone: (306) 953-8540
FAX: (306) 953-8573

Telephone orders will be accepted but they must be confirmed by FAX or mail before production will begin.

The request should include:
- satellite and sensor,
- date of preferred coverage - from/to,
- maximum cloud-cover acceptable,
- product required,
- brief description of user’s application,
- location - Landsat WRS path and row and/or latitude and longitude coordinates.

NOTE: User may preselect the scene required using CCRS catalogs, microfiche, or "QUERY" system.

The User Assistance and Marketing Unit, CCRS, Ottawa is the contact point for all requests for general information and applications oriented advice.

Telephone: (613) 952-2717
FAX: (613) 952-7353

Private Sector Involvement
An important part of the CCRS mandate is to assist the remote sensing industry in Canada. From the very beginning the Centre’s operational style has been industry oriented. A wide variety of contractual arrangements for the development of required systems and the provision of both satellite and airborne data and services has met the needs of the Canadian user community and served as a means of technology transfer. A continuing example of this style is the operation of the CCRS satellite receiving stations reception, archiving, and production systems by Canadian industry from 1972 through to to-day. A new dimension to the industrially oriented operations of the Centre will begin in 1990 with the commercialisation of data production, distribution and marketing in Canada for Landsat and SPOT data. The Government (through CCRS) will continue to be responsible for the reception and archiving activity and for the maintenance of the archived data catalogue. Industry will be completely responsible for responding to the needs of the user community for data products. Industry will be loaned the Canadian MOSAICS production system and will have complete financial responsibility and complete freedom in both market and product development. The program initially involves data from Landsat and SPOT, however, the plan is to expand the industrial responsibility for commercial opportunities to other satellite programs, including both ERS-1 and, of course, RADARSAT. It is felt that this concept ensures the continuity of the National commitment to remote sensing by having the Government responsible for the maintenance of the data archive while at the same time providing industry with the means and incentive to develop the continued use of remote sensing as a tool for resource and environmental management.

RADARSAT
Canada’s first Earth Observation satellite received final approval on September 13, 1989 when the Minister of Industry, Science and Technology Canada signed a Memorandum of Understanding with nine provincial partners on participation in the Canadian led international RADARSAT project. Also, a Memorandum of Understanding outlining the responsibilities and cost-sharing arrangements of international partners to this project has been negotiated with the United States National Aeronautics and Space Administration (NASA) and the National Oceanic and Atmospheric Administration (NOAA) of the United States Department of Commerce. To promote the global use of RADARSAT data, the federal government has also negotiated a Memorandum of Understanding with a new Canadian company called RADARSAT International. Apart from the data rights retained by the Canadian and U.S. governments, this company will be assigned the sole rights to market all RADARSAT data internationally.

RADARSAT will be launched on a medium-class NASA rocket into a circular, sun-synchronous polar orbit. The satellite will circle the earth at an altitude of 800 kilometres and will complete approximately 14.4 orbits during a 24-hour period at a rate of 100 minutes per orbit. The satellite will cross the same point at the equator every 16 days and cover the Arctic daily. RADARSAT is planned to operate in space for a minimum of five years. The payload module of the satellite consists of the synthetic aperture radar (SAR), whose antenna will be 15 metres long and 1.5 metres wide when unfolded in space.

Compared to conventional radar satellites, the SAR will be versatile and innovative. It will provide a great deal of flexibility in the amount of coverage, the angle of observation, the ability to zoom in for higher level of detail for a smaller area or out for less detail of a larger area. The antenna will be able to point its beam anywhere within a swath of 500 kilometres between 20 and 50 degrees off the satellite’s orbital path.

RADARSAT will provide valuable economic and scientific information on ice conditions, forests, and geological formations. The satellite’s ability to collect data through darkness and cloud will be particularly useful in penetrating the almost constant cloud cover of the equatorial rain forests and Canada’s coastal regions.
Data Issues and Remote Sensing

Session F2
Survey Control

Developing the Foundation: A Geographic Coordinate Data Base for Forest Resource Planning

High Altitude Laser Profiling For Vertical Control in Small and Large Scale Mapping.

An Introduction to the Global Positioning System and its Use in Ground Truthing Spot Satellite Imagery
Developing the Foundation: A Geographic Coordinate Database for Forest Resource Planning

Lisa Crouser
GIS Specialist

Larry Guntrip
Program Director LiS
Infotec Development, Inc.
500 NE Multnomah, Suite 352
Portland, Oregon
97232
USA

Abstract
The U.S. Department of Interior, Bureau of Land Management, Oregon State Office is in the final stages of developing a 7 000 000 acre data base to support natural resource management planning efforts for the 1990s. Crucial to this effort has been the development of a Geographic Coordinate Data Base (GCDB) utilizing Public Land Survey Coordinate Computational System (PCCS) software. The process involves abstracting cadastral survey information and computing 'as surveyed' coordinates to serve as the base registration for the remainder of the data base. This paper will provide an overview of the process, a view of the impact on GIS processes, and the future development of the GCDB nation-wide by BLM and Infotec.
Introduction
Infotec Development Inc. has been the GIS/ADP computer support services contractor for the BLM Oregon State Office for the WODDB project since 1986. Infotec has also been awarded two contracts by BLM for development of GCDB data on a national scale. This paper approaches the subject of GCDB development from the perspective of a support services contractor.

The Western Oregon Digital Data Base (WODDB) is a U.S. Department of Interior, Bureau of Land Management (BLM) project to develop a digital data base to support Resource Management Plans (RMPs) scheduled for completion in the early 1990s on BLM lands in western Oregon. The project has involved the digital mapping of over 7,000,000 acres at a scale of 1:4,800 and now includes a substantial number of thematic layers.

The accomplishment of the project in the very tight time frames required the development of a sophisticated tool box of hardware, software, and personnel. One of the key tools was processing and software to link the data base to the "real world" in terms of cadastral survey accuracy. This was necessary to alleviate inaccuracies in existing map products, and to avoid being required to digitize geo-reference points repeatedly throughout the project.

Background
Historically, the basis for most of the digital data utilized in BLM GIS data bases has been derived from existing maps. Most of these maps were old, intended for other purposes, and in most cases, simply "cartoons" of the real world. The WODDB project started with new aerial photography, the best available established geodetic control, and a scale of mapping that could provide a tremendous level of detail for the field user. The problem was to register this mass of digital information accurately to the surveyed world.

PLSS
In order to understand the complexity of the problem, a basic understanding of the U.S. National Public Land Survey System is necessary.

A. Basics of the PLSS Today
Starting with the Land Ordinance of May 20, 1785, which provided for surveys of public lands within the state of Ohio, a variety of acts, ordinances and manuals of instruction, accompanied by changes in governmental responsibilities, resulted in a very large and complex Land Survey System. There are various kinds of projection data systems, different storage methods and various data recording procedures, including computerized, micro-fiche, or simply a copy of an original linen. Every agency or entity has its own methods or priorities depending on the need, funding or political concerns. Thus, surveying data is scattered, not standardized and at times, out of data.

Whatever the form of the data, certain basic features are common in the PLSS, including:
1. Independent Initial Point,
2. Base Line,
3. Standard Parallels,
4. Meridians,
5. Townships,
6. Sections,
7. Subdivision of Section,
8. Irregular Tracts, and

B. Kinds and Sources of PLSS Data
PLSS data may be obtained from a wide variety of sources, including at least those described below.

1. Cadastral Survey Plats and Field Notes
The primary and major component of the PLSS is cadastral survey plats and field notes. These plats and field notes are stored mostly on microfiche at state BLM offices. Surveys of some military reservations and some Indian reservations that were subdivided for Indian allotments have not been put on microfiche and are available from hard copies only.

2. Rectangular Surveys
The most common type of cadastral survey plats and field notes is the rectangular survey. The majority of the western states have rectangular land surveys. In the states covered, there are 26 initial points of the rectangular survey. From these initial points, locally named principal meridians and baselines were established. Townships 6 miles square were laid out and referenced to the principal meridian and baselines.

3. Special Surveys
Special surveys involve unusual applications or departures from the rectangular system, and are of at least six kinds:
A. Tract and Lot Survey,
B. Townsite Survey,
C. Small Tract Survey,
D. Metes and Bounds Survey,
E. Mineral Survey,
F. Millsite Survey.

USGS topographic maps show improvements that have been placed upon the ground and may be referenced to the PLSS corners. In areas where blunders and gross discrepancies exist, these maps can aid in determining where the error should be left. Also, some of the PLSS corners have been identified on photographs, and their locations are sometimes shown on the USGS topographic maps.

5. Control Diagrams
USGS and NGS Control Diagrams and Lists give the location and geographic coordinates of the main government geographic control points established in the field.

6. Status Maps
These maps provide the status of all federally owned land. Also shown are mineral rights, mineral leases, mining
claims and road, powerline, pipeline and other utility rights-of-way.

7. Protraction Diagrams
Protraction diagrams present plans for extension of the rectangular system over unsurveyed public lands, based upon computed values for the corner positions. These diagrams, in the lower 48 states, are commonly used to identify mineral leases and mining claims in unsurveyed lands.

8. Connection Sheets (Mining Districts)
Connection Sheets show all the mining claims in a mining district. In some areas, supplemental sheets are available to show all the mining claim surveys. From these sheets, one can determine the approximate locations of mining claims and the names and survey numbers of mining claim surveys. Connection sheets are indexed by mining district.

9. Aerial Photos
Aerial photos are available for most areas from government agencies and USGS. In others, state, county, local governments and private sources have complete or partial coverages of the area. From photos, one can determine land use patterns which can help in identifying where survey blunders may have occurred in the PLSS.

10. Survey Records (state, county, private, etc.)
Many records of PLSS corners were created by state, county, private individuals, utility companies, etc. Utilities, while performing their field surveys, tied into PLSS corners along the route of the utility line in order that legal descriptions could be written for the line and access descriptions. Many of these surveys have already been tied into the State Plane Coordinate System and have field geographic coordinates on them.

All types of private surveys have been performed on PLSS corners but the availability of the records varies in different states and localities. These records are found mostly on mylar or paper, and are referenced in many different ways. Most indexes are referenced to the index system in use by the federal government for the PLSS, but modified to suit their individual needs.

11. Right-of-Way Descriptions
Many right-of-way descriptions have been written for access and utility purposes. With different degrees of accuracy, these descriptions have been tied into the PLSS corners. In many cases by using aerial photos, 7 1/2 minute USGS series maps, etc., the geographic location of the PLSS corners called for in the right-of-way description can be determined. Right-of-way descriptions are indexed on the master title plats of federal lands.

12. Special Situations (Eastern States)
In the states of Ohio and Indiana, there are eight public land rectangular surveys with no initial point as an origin for townships and range numbers. These surveys are indexed by survey name. They were the surveys used to structure the present day rectangular system of surveys. No rectangular surveys were performed by the federal government in the colonial states. In these states, metes and bounds surveys were privately performed. Terrain and physical features called out in legal descriptions is what mainly controlled these surveys. These private metes and bounds surveys are filed by many different indexing systems in the different counties.

PCCS
As can be seen, the PLSS is a very complex source of information. Using any digitizing process for data base construction from most of these sources will make it very difficult to achieve any quality or reliability. To address this problem, Mr. Bill Bail, BLM Denver Service Center, developed the Public Land Survey System Coordinate Computational System (PCCS) which was brought to Oregon in 1986 as a Beta test. The software was so successful that it was adapted as part of the WODDB effort. By the end of 1988, 377 townships had been processed through PCS for the WODDB effort and entered into a data base (GCDB).

A. Methodologies
Abstraction, data input/data entry, digitizing and computation are the major procedural categories within the GCDB development process. The sections immediately below discuss each of the procedural categories with special emphasis on how they relate to the achievement of a high quality, accurate and complete GCDB.

1. Data Abstraction
Abstraction involves the translation of surveyor notes, plats, maps, diagrams and other survey information into a listing of bearings and distances which define the surveyed locations of PLSS corners in a township. The process of abstracting data is very exact. The first step involves identification of all source documents available for a given township. Next, the type of control in the township is determined. Third, data for adjacent townships are examined to determine which township contains the most recent (or best) survey data for the township boundary. Finally, all source documents for the township are examined to identify the one from which PLSS corner information will be abstracted. Only after all documents have been examined and those of the highest quality have been identified, does the surveyor begin to abstract information.

The township boundary is abstracted first, followed by the section lines. If irregular lotting exists, these are abstracted only after the surveyor again examines the survey notes to verify how the lotting was accomplished. Any special surveys (small holding claims, homestead entry surveys, U.S. surveys, townsite surveys, tracts, etc.), grant boundaries, mileposts, mineral surveys and meanders are also abstracted. For each corner abstracted, the surveyor records bearing, quadrant, distance, elevation, control, type of survey line, source, origin and average terrain elevation on the appropriate abstract form.

If there are many non-rectangular surveys in a given township, the PLSS corners from the non-rectangular surveys may be captured by digitizing, rather than abstraction and calculation.
During the abstraction process each corner abstracted is identified with a unique point identification code. The point ID code is a mathematically logical system for identifying all points in a township. The origin is taken as the SW corner of the township, and assigned an ID of 100100. This six digit number consists of two parts. The first three digits define the distance east of the west boundary of the township, while the last three digits define the distance north of the south boundary. Each section line is represented by an even 100 number. Points intermediate to section lines are assigned numbers based on the number of chains north and east of the section lines. All non-rectangular survey corners are assigned point IDs such that the first three digits indicate the type of survey. The last three digits are used for sequencing of the survey lines. After all PLSS corners have been abstracted, the parentheticals for the township are determined. Parentheticals are usually 1/16 corner points that will be computed by the PCCS programs based on the locations of 1/4 corners. These parentheticals are marked on the survey plat and entered on the Section Abstract Sheet. Next the geographic center of the township is determined. This is a reference point that is utilized by several of the PCCS programs to compute tangent plane coordinates and geographic coordinates. The township center or origin is computed in most cases by adding 3 minutes latitude and 3 minutes longitude to the coordinates of the SW corner of the township. Since this point is for reference only (i.e., like a ‘seed’), it need not be known precisely. The origins for townships which are not close to square will need to be determined by the surveyor and recorded on the abstraction forms.

2. Data Entry
Data entry involves input of data from abstraction sheets into a computer. Each of the PCCS data entry programs uses a prompted entry system, i.e., the operator is prompted for each bit of information required by the program. Two pathways have been developed for initial data entry. The first requires a surveyor to input data from abstraction sheets using a direct data entry program. The second utilizes a data entry person, at lower cost, to input abstracted data utilizing a raw data entry program. A surveyor will later review and organize the raw data entered by the data entry person. Data is entered into the PCCS system in a specific order. Every program prompts for the township, range, meridian, and state first. Next, the point ID, coordinates and elevation of any control points (at least two are required) located in the township are entered. All bearings are entered and stored in compressed degree-minute-second format, which are expanded automatically prior to computations by the PCCS programs. This part of the file is referred to as the control section. Standard rectangular township survey data may be entered using any of the following programs: raw data entry, direct data entry, or control station coordinate file.

The survey measurement data subsection begins with entry of the township boundary data. This consists of the point ID for the starting point, point ID for ending point (TO/FROM), the bearing, quadrant, distance and elevation (if known). Boundary lines are entered starting with the eastern boundary, entering data from south to north. The remaining south to north section lines are then entered, working to the west, followed by the east to west section lines, the most southerly line entered first.

Non-standard rectangular portions of townships and special surveys are entered separately using a different program. In this program, initial data entry is identical to that required by standard townships: prompts for township, range, meridian, and state, followed by entry of control point data. The order of input of survey measurement data is up to the discretion of the data entry person. The order will usually be determined by the surveyor as they abstract the data initially.

3. Digitizing
Capture of PLSS corner coordinates using the digitizing process requires the use of a digitizing tablet and the computer editor software. Digitized coordinates are incorporated into the GCDB database only under the following two conditions: coordinates are needed from positively identified PLSS corners from USGS Quadrangle Maps such as 7.5 minute, or 15 minute quads, coordinates are also needed from PLSS corners shown on accurately drafted BLM Cadastral Survey Plats, where the time required to calculate corner coordinates using PCCS programs would greatly exceed digitizing time. Digitizing is usually employed to capture corners from small parcels of land such as mining claims, small survey tracts, or small with holding claims.

4. Computations
The computation part of the GCDB project involves selection and use of the PCCS programs for evaluation and adjustment, creation of PGCF files, subdivision of sections, graphics (PGCO), interfacing PCCS to other programs, and creation and maintenance of the PCCS database. PCCS programs are simply analytical procedures. The computation programs are designed to not introduce mathematical error into the process. Therefore, coordinate results of PCCS programs reflect the inherent accuracy of the original survey data abstracted from plats or survey notes. The knowledge and skill of an experienced cadastral surveyor is required to properly process data through the PCCS system. An experienced surveyor must analyze the data and determine which analytical computations should be performed to properly process the data.

5. Evaluation and Adjustment
There are two types of evaluation and adjustment programs, those which provide evaluation only, and others which will adjust the data and provide evaluation of the adjustment process. The main options are: the use of previously computed coordinates (usually township boundary) as fixed reference points for adjustment of interior corners, selection of network adjustments (all points in a township) or adjustment of an individual traverse line, and how extensive a report is required (original data + adjusted data + adjustment results or just adjustment results).

B. Interfaces
Programs are available for translating PCCS PGCF-type files into ADS (BLM’s digitizing software), ALMRS (BLM’s
Land Records system), and AUTOCAD formats. To transfer the files into ADS, a program must be run to create a file with Universal Transverse Mercator (UTM) coordinates. To create an ALMRS file, two programs must be run on a PGCF file. First, a program which creates a list of non-standard PCS point identifiers that is used by a second program to create the ALMRS Point Identifier Translation file. All three interfaces are one-way. At the present time, AutoCAD, ADS or ALMRS files cannot be imported into the PCS data base.

The interface to AutoCAD was utilized to accomplish WODDB using a grid generator built into AutoCAD to construct the BLM ownership theme to the 1/16 corner level. The key point, however, is that the coordinate points generated in the PCS program were used to geo-reference all of the data generated in the project. The data was then translated into MOSS (BLM’s GIS software) for analysis purposes.

For the first time in the agency’s GIS activities, the digital data is truly registered to the field located brass caps, to a surveyor’s level of accuracy.

Future
The Bureau of Land Management is committed to developing a full Land Information System (LIS) to manage the lands under the agency’s jurisdiction. Their definition of LIS is that it is composed of three parts: Resource Data; Land Records; and the National Geographic Coordinate Data Base (GCDB). They are collecting Resource Data as needed; collecting ALMRS records in digital format both in-house and by contract; and have let two contracts to Infotec Development, Inc. to abstract and process GCDB data for the continental U.S. and Alaska over the next four years. This data will be used by BLM to populate the National GCDB.

The final BLM GCDB will be the primary source for coordinate data for most GIS data bases, and as such, will contribute substantially to the quality of the data used for resource management. The first major effort to expand the use of the GCDB outside of BLM is now beginning in Oregon with a pilot project under the Northwest Land Information Systems Network. This project will involve two counties, Benton and Harney, in a data sharing and development process based on the GCDB developed in Oregon. The success of this project will probably lead to further work in the Oregon-Washington region.

References


Disclaimer
Reference herein to any specific product, process, or service by trade name, trademark, manufacturer, or otherwise does not constitute or imply its endorsement by the United States Government or the Department of Interior, Bureau of Land Management. Questions relating to the WODDB project should be directed to the Public Affairs Office, Oregon State Office, Bureau of Land Management, Portland, Oregon 97232.
High Altitude Laser Profiling for Vertical Control in Small and Large Scale Mapping

J. Blair, B. Hlasny, J. Schleppe
McElhanney Geosurveys Ltd.
780 Beatty Street
Vancouver, British Columbia
V6B 2M1
Canada

Abstract
Airborne Profile Recording systems for large and small scale mapping have now been replaced by laser technology combining integrated components which include bore-sighted standard cartographic cameras and small tracking cameras, precise altimetry and GPS positioning. McElhanney has recently completed a contract covering the islands of Kalimantan and Sulawesi in Indonesia providing vertical control and aerotriangulation for 1:50,000 map production. This contract required the installation of a high altitude laser system mounted onboard a fixed wing aircraft, to provide photo-identifiable vertical points which were transferred to mapping photography. The laser system, designed by Holometrix of Cambridge, Massachusetts, is the first privately owned high altitude unit to be integrated and designed specifically for such mapping applications. A total of 11,000 line kilometres were flown in four months over heavy jungle terrain providing true ground data to inaccessible areas. Project field operations, profile results and vertical transfer methodology are presented in this paper.
Introduction

Airborne profile recording (APR) for topographic surveys were first undertaken in the late 1940's with particular application in Canada to the vast northern wilderness areas. The early systems were timed radar measurements, using 1 1/2 meter diameter antennas and provided profile accuracies of approximately 10 meters. These systems were used extensively on projects throughout the world for map productions at small scales such as national 1:50 000 series.

Recently with the advent of laser technology, more accurate systems have been developed, and APR is being used not only in mapping applications, but also in hydrographic surveys, route profiling, forest inventory and water quality sensing.

McElhanney's development of an integrated laser system came from the Company's involvement in a large survey and mapping program in Indonesia. This program is being carried out as the Resource Evaluation Aerial Photography Project (REAP) and was designed to set up a physical and institutional capability in Indonesia to handle an accelerated program of natural resource development, particularly forestry, mining and oil and gas exploration. Canadian aid has sponsored the project over 3 phases commencing in 1979. Phase I was the establishment of horizontal and vertical control points using Doppler satellite observations to form a base for 1:50 000 mapping. This was completed by McElhanney in 1981 (A total of 238 stations established covering all of Kalimantan, Sulawesi and some of the southern islands). Phase 2 provided aerial photography coverage and Phase 3 has provided densification of vertical control using the laser APR, for aerotriangulation and numerical adjustment of the final map compilation.

The APR for use on Phase 3 was developed to provide high altitude terrain profile data at an economical cost using fixed wing aircraft. The system was developed for portability and to operate in relatively small aircraft with standard aerial camera ports. It is designed to acquire ground profiles for applications such as route profiling, cross-sectioning and vertical control for topographical mapping.

APR Profiling Principles

APR equipment when mounted and operated in an aircraft (airplane or helicopter) measures vertical profiles of terrain with respect to a reference isobaric surface.

The aircraft flies along a constant pressure altitude anywhere from 1 000 to 30 000 feet above the ground level. Pulses of laser energy (travelling at the speed of light, 186 230 mi/sec) are transmitted in a narrow beam and the elapsed time between transmission and return is a measure of the distance between the aircraft and the ground. These measurements are reproduced as a continuous line or profile of a narrow strip of ground traversed by the aircraft. A Barometric Reference Unit (BRU) senses deviations of the aircraft either up or down from a constant pressure level. The flight path of the aircraft is recorded by a tracking camera (70 mm frame or continuous strip camera), or other onboard navigation system.

After suitable adjustments to remove pitch, roll, BRU deviation and isobaric surface slope, and to adjust the terrain profile to ground datum, the profile is used to provide vertical control for mapping or other DEM applications.

The general method by which heights are obtained from profiles is as follows. The height of the reference isobaric surface must be determined at the start and end of the flight line by profiling over terrain of known height above the selected datum (i.e., sea level). The height of the reference isobaric surface over any unknown terrain point is interpolated from each end with respect to time. Knowing the aircraft to ground distance (from the laser range) over the unknown terrain point and the height of the reference isobaric surface, the height of a point on the terrain (APR point) may be calculated. By transferring selected points on the profile track from 70mm tracking film to the mapping photography, heights of APR points may be obtained on the mapping photography along the profile track or input by coordinates from an onboard navigation system.

APR System Components

The McElhanney airborne laser system consists of five separate sensors, four of which are interfaced to the central processing unit and each drawing power from the central power distribution module. The system is essentially a passive data collector with the only control being over the tracking camera. The central computer waits for information from the altimeter, pitch/roll meter and the laser. It in turn triggers the tracking camera shutter and advances it based on operator inputs. It stamps all information coming in with a time tag, allowing it to later reassemble and correlate the data at the post processing stage. A description of each component is given below. Figure 1 illustrates the interfacing of the components.

Computer and Peripherals

The central processing unit for the laser system is the McElhanney designed M CAT.DC powered navigation computer which was originally developed for the Canadian Hydrographic Service for airborne and hydrographic survey applications.

The computer itself is composed of four major components. The processing unit is a PC - AT compatible mother board utilizing a 80286-10 high speed 16-bit microprocessor chip and a 80287 math co-processor chip. It has built into it three RS-232 serial ports and two parallel ports, which are used to integrate the survey sensors of the laser system. The M CAT's power supply is rated at 19-36 volts and is polarity protected allowing it to function in an airborne environment with minimal power concerns. The mass storage devices for the M CAT include a 20 Mbyte hard drive, a 1.2 Mbyte, 5.25 inch floppy drive and a 720 M byte, 3.5 inch micro-floppy drive. The system display monitor is an ASK LCD flat screen and graphics adaptor. The screen is lightweight and requires no power input other than the power drawn by the graphics adaptor card in the computer. The entire computer is designed to be rack mountable.
Laser Profiler
The laser used by the McElhanney system is the ACCI PRAM IV Model LRY-1000 pulsed laser, which was designed specifically for high altitude airborne service. It has a modular design consisting of a PRAM IV control console and a Litton NT-90 flash lamp-pumped Nd:YAG near-infrared laser transceiver. The transceiver utilizes side by side optics in a compact package designed for easy mounting in an aircraft. The separate control console is designed to be rack mountable. The laser has a peak power rating of 114 mJoules and can operate at speeds of 5, 10 and 20 pulses per second at ranges up to 6000 meters. Range accuracy of the laser is 0.1 meter. For the REAP III project the laser was always operated at 20 pps. As a safety precaution, the laser is fitted with filters to improve the eye-safe range to 400 metres.

Power to the laser transceiver is supplied directly from the aircraft, whereas, power to the PRAM IV console (28 VDC) is supplied via the McElhanney central power distribution unit. The PRAM IV profiler is interfaced to the M CAT computer via an RS-232 port. The PRAM IV profiler sends a steady stream of range and system status information at a rate of 20 ranges per second to the computer, which collects the data in a software-created ring-buffer.

Digital Elevation Computer
The laser system utilizes a Davidson Elevation Computer Model D101 as a barometric pressure altitude reference. This solid state unit incorporates pressure and temperature sensors in a software controlled unit. The pressure sensor is maintained in a controlled temperature environment using an operator-controlled oven. The pressure sensor is connected to the aircraft's static system. Vertical resolution is 0.1 meter.

The D101 is built into McElhanney's DC power distribution unit and power (12 VDC) is supplied from the distribution module. The D101 is interfaced to the M CAT via an RS-232 port. The D101 measures the pressure once every 5 seconds and sends the pressure and temperature information to the computer.

Tracking Camera
The system's tracking camera is a Hasselblad 500 ELM 70 mm format camera complete with an autowinder, a 50 mm super wide angle lens and a detachable 120 exposure film magazine. The film magazine is equipped with a built in DA-1 serial interface and a detachable DE-32 Datapac which allows the system to automatically annotate the 70 mm tracking film with data, time, roll number and frame number.
The camera is triggered via one of the central processing computer's parallel ports which sends a command to the McElhanney-built triggering device which translates the command into a triggering pulse which is in turn sent to the camera through the power winder power cable. Programming the DE-32 datapac for film annotation is accomplished from the computer via one of the serial ports and is normally done prior to the start of each line. The film utilized by the tracking camera is Kodak's 2402 Plus-X aerographic film. The frame interval was controlled by the central computer based on the operator selected ground speed, flying height and forward overlap. A forward overlap of 60 percent was used for the entire project.

**Pitch and Roll Meter**
The MDL TRIM Cube Model 40S provided pitch and roll information for the laser system. It is a two axis inclinometer using electrolytic gravity sensors with a single chip microcomputer used to demodulate the sensor information and produce a serial data stream. Roll and pitch information is sent to the computer 5 times per second, where it is recorded with the other sensor information. The meter has a range of 40 degrees and a stationary accuracy of 0.25 degrees. In the airborne environment with its accompanying accelerations, the accuracy is severely decreased. Power to the TRIM cube (12V DC) is provided by the central power distribution module.

**Power Distribution Module**
The aircraft power (28V DC) is fed into the central power distribution unit, which then distributes it to the four survey sensors. Power at 28 volts DC is provided to the PRAM IV processing console, and the M CAT computer. The power supplied by the aircraft is further broken down to 12 volts DC and 6 volts DC. The 12 volt DC power is supplied to the digital elevation computer and the 6 volts DC is supplied to the tracking camera and datapac.

**System Installation**
The central processing unit, laser processing unit, digital elevation computer and power distribution module are all mounted in one rack. The remaining survey sensors, consisting of the laser, pitch and roll meter, and 70 mm tracking camera are mounted on a heavy duty aluminum alloy mounting base designed to fit into an RC10 standard camera mount. Figure 2 illustrates the mounting arrangement in the aircraft.

**Large Format Camera**
A large format camera was used during the REAP III project to supply supplemental photography. The Zeiss RMK A 8.5/23 camera complete with two magazines, 8.53 cm lens, drift sight and IRU were employed by the McElhanney system. The large format camera was not interfaced to the central computer system, although synchronization is possible using recorded times. Photography was captured with a forward overlap of 60 percent using Kodak 2405 Double-X Aerographic film.

**Post Processing Peripherals**
For post-processing the APR data, the M CAT computer is interfaced to a Okidata Microline 320 high-speed dot matrix graphics printer and to a Houston Instruments Hipad Plus Model 9012 digitizing tablet and cursor. The printer is interfaced to one of the M CAT's parallel ports while the digitizing table is interfaced to one of the computer's serial RS-232 ports.

**Data Collection Software**
Program MAPR V1.0 is used by the McElhanney system for the real-time collection of APR data. MAPR V1.0 was written by McElhanney in the C language and uses serial interrupts extensively to receive data from the survey sensors. The program is fully interactive and menu-driven, allowing the operator access to most levels of the program while the data collection is underway.

As data is read by the program, it tags the data with a time tag and then stores it in a unique device file on the hard disk. There is a file created for each of the devices, which include the laser, altimeter, camera and pitch/roll meter. The time tags allow the post processing section of the program to later reassemble the data and correlate it for adjustment and profile display.

The operator has several displays to choose from while operating in the real-time environment. The first display shows the status of all of the devices interfaced to the computer along with selected operator input parameters which control the camera exposure interval. Also displayed is time-to-exposure for the tracking camera and an estimate of the amount of film used by the tracking camera. A second display which can be seen simultaneously with the first shows some of the laser ranges, all of the altimeter readings and all of the tracking camera triggering pulses. Other displays include an error display showing all errors incurred by the system and a buffer display showing the status of internal software device buffers.

**APR Data Processing**
Post-processing of the data collected is accomplished using McElhanney-developed software packages designed to run on PC compatible computers. The following sections explain the concepts behind the profiling package and summarize the features.

The data acquisition system is essentially a passive data collector, exerting control over the tracking camera only. It tags all data coming into it with a time tag and then stores the data in a device specific file. For the REAP III project there were normally four files created containing the time tags in addition to: laser data; altitude data; frame centre data; pitch and roll data. The files are stored on floppy disk and accessed for adjustment and profile generation. All data collection, adjustment and profiling is performed with program MAPR V1.0. Processing flowchart is shown as Figure 3.
Examination of Profiles
At the end of each day's flying, all data collected was inspected for completeness. The data was processed through program MAPR and the unadjusted profiles were inspected for anomalies; poor laser tracking of water surfaces (necessary for profile adjustment); large gaps in data; improper exposure and firing of tracking camera. The data examination was typically completed within 24 hours.

APR Data Adjustment
The isobaric surface in the area of operations was quite stable, since it was close to the equator. This eliminated the need to apply an isobaric surface drift angle correction to the profiles. Instead, a linear correction was determined for each profile which corrected for the line's initial datum error and the elevation discrepancy at the end of the line. Two components were computed for each line using an adjustment routine resident in MAPR. The operator would inspect the unadjusted profile and input the time, raw elevation and the controlling elevation for the control points or sea level at the start of the line and end of the line or APR traverse. The program then automatically computes a translation component in metres which corrects for the line's initial datum error and a scaling component in metres per second which accounts for the residual discrepancy calculated at the end of the line or traverse. In addition, the operator can enter intermediate points into the adjustment and it will give misclosures for these points along with intermediate correction values. The operator is then free to interpret the results and enter the values most suitable as corrections for later profiling of the APR data. A printout of each profile adjustment is made and logged along with a graphical representation of the APR lines or traverses and the corrections. Corrections entered by the operator for each line, if satisfactory, were then saved in an initialization file which is stored for each line and can later be recalled when the data is to be inspected. This eliminates the need to re-enter the correction components and reduces the possibility of operator error.

Profile Display
The McElhanney APR data collected during the project was graphically displayed for interpretation using program MAPR. MAPR has a high degree of display flexibility permitted by the concept of time tagging. The display modes are summarized below.

Display By Frame: Here the profile displayed is bounded on the horizontal axis by the centres of two of the tracking film frames. All profile information between the two frame centres is displayed. The user has the option to display the terrain, the aircraft's altitude or both in addition to the roll and pitch meter readings.

Display By Time: Here the profile displayed is bounded on the horizontal axis by two times selected by the operator. All APR data is displayed between these two times. This allows the operator to generate a profile with any horizontal scale.
APR (MAPR) PROCESSING FLOWCHART

Figure 3
Table 1. Summary Of Linear Adjustment Data And Corrections For
MAPR - McElhanney Aerial Profile Recorder - K1 Block - REAP III

<table>
<thead>
<tr>
<th>Line</th>
<th>Correct Elev (m)</th>
<th>Observed Elev (m)</th>
<th>Time (hh:mm:ss)</th>
<th>Translation (m)</th>
<th>Scale (m/s)</th>
<th>Misclosure (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>116V6</td>
<td>0.7</td>
<td>0.6</td>
<td>08:18:56.98</td>
<td>0.1</td>
<td>0.00010</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>-0.3</td>
<td>09:42:32.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>116_SL</td>
<td>0.0</td>
<td>23.0</td>
<td>08:05:49.95</td>
<td>-23.0</td>
<td>0.00037</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>21.0</td>
<td>09:36:20.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>116V5</td>
<td>2.0</td>
<td>7.8</td>
<td>08:16:22.25</td>
<td>-5.8</td>
<td>0.00022</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>4.2</td>
<td>10:18:55.79</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>116V6</td>
<td>0.7</td>
<td>0.6</td>
<td>08:18:56.98</td>
<td>0.1</td>
<td>0.00010</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>-0.3</td>
<td>09:42:32.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>116_SL</td>
<td>0.0</td>
<td>23.0</td>
<td>08:05:49.95</td>
<td>-23.0</td>
<td>0.00037</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>21.0</td>
<td>09:36:20.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>116_SL2</td>
<td>0.0</td>
<td>23.0</td>
<td>08:05:49.95</td>
<td>-23.0</td>
<td>0.00037</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>21.0</td>
<td>09:36:20.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>117V2</td>
<td>1.5</td>
<td>13.9</td>
<td>09:41:54.54</td>
<td>-12.4</td>
<td>-0.00024</td>
<td>-0.6</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>13.5</td>
<td>10:23:17.23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>117_SL</td>
<td>1.5</td>
<td>13.9</td>
<td>09:41:54.54</td>
<td>-12.4</td>
<td>-0.00024</td>
<td>-0.6</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>13.5</td>
<td>10:23:17.23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>118V1</td>
<td>0.0</td>
<td>-1.0</td>
<td>08:43:15.00</td>
<td>1.0</td>
<td>0.00003</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>-1.1</td>
<td>09:49:46.50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>118V2</td>
<td>0.0</td>
<td>-7.3</td>
<td>07:55:21.00</td>
<td>7.3</td>
<td>0.00026</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>-8.5</td>
<td>09:13:42.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>118V3</td>
<td>17.2</td>
<td>32.4</td>
<td>09:19:42.71</td>
<td>-15.2</td>
<td>0.00210</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>13.9</td>
<td>09:41:54.54</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

When a profile is created, MAPR correlates the laser, altimeter, pitch and roll readings as well as the frame centres using their time tags. Since the system records some 100 laser pulse within the 5 seconds update time of the digital elevation computer, the program does a straight line interpolation of the altimeter data based on the time of each laser pulse.

An operator-controlled filter helps eliminate spurious data within the profiles. An indicator at the top of the profile shows where data has been rejected.

To facilitate point selection along the profile the program utilizes a digitizing tablet and cursor. The operator digitizes a left and right photo centre on the tracking film, after which he can digitally display elevations, altitude, roll and pitch along the profile as the cursor is moved within the tracking film. This, coupled with the stereo images on the tracking film, allows the operator to identify features and transfer them to the mapping photography for mensuration. Selected points were also entered automatically into a log file by pressing one of the cursor keys.

Printouts of the profiles are available using a high-speed dot matrix printer. The operator has the option to print one profile only or the entire line (at any scale). Figure 4 is a reduced copy of a printout.

APR Project Application
The Indonesian project provided a real test for both hardware and system capability. Terrain in Kalimantan and Sulawesi varied from low level swamps and jungles to densely forested mountains, with temperatures consistently in the low 30's and high humidity. Elevation varied from sea level to more than 3000m.

Field base camps at air strips were established throughout the project area and general transportation was either by aircraft or boat due to the lack of roads.

Weather was the major contributing factor in profile production. Generally, flight missions were achieved during the morning before any cloud build-up developed. Extreme heat, haze and cloud affects laser penetration and restricts identification on aerial photography.
Flight profile lines were chosen to meet mapping specifications for vertical point density and also to follow the geography of the land so that major features such as mountain ranges, rivers, lakes and coastal areas were easily identifiable. Daily flight planning was critical to ensure planned flight paths were covered and profile gaps were fully completed.

Profile lines commenced and terminated on points of known elevation such as sea level, lakes, rivers or large open areas. Additional check and common points were tied in along the flight path to improve closure accuracy and detect any blunders. After each day’s mission, all profiles were verified and adjusted prior to the next day’s mission planning.

Profile data and aerial photography were forwarded to photogrammetrists based in Jakarta. Here, the vertical point selection, identification on the 70mm film and transfer to the mapping photo was completed. The results were processed through the aerotriangulation adjustment program PATM which verified the elevation value with the identified point selected.

Profile Results
Flight missions were run from a half hour to 8 hours per day (averaging 3 hours over the project) depending on the weather conditions. A sample printout of line closures is shown in Figure 4.

The misclosure of the flight line is calculated on a meter per second basis for the total flight mission, and applied to the laser ranges on a linear time basis.

Elevation closures were 0.5 m to 2.0 m depending on the flight time, which relates to a 2 to 6 cm vertical correction per kilometre.

Figure 4. Real Time Profile Printout
An Introduction to the Global Positioning System and Its Use in Ground Truthing Spot Satellite Imagery

Arthur F. Lange
Trimble Navigation, Ltd.
585 N. Mary Avenue
Sunnyvale, California
94086
USA

John Stenberg
National Park Service
South Florida Research Center
Everglades National Park
P. O. Box 279
Homestead, Florida
33030
USA

Abstract
The NAVSTAR Global Positioning System (GPS) provides accurate position data which can be used to aid the users of satellite imagery and other remote sensed data. When fully deployed, GPS will use 24 NAVSTAR satellites in a constellation arranged to provide 24-hour coverage with 3-dimensional position fixes accurate to better than 25 meters. In differential mode, the accuracy will be better than 5 meters. For high accuracy survey applications, the errors will be in the millimeter range. The GIS applications of GPS include first geo-referencing satellite imagery and then using the GPS receiver to navigate to a site for ground truthing the satellite image. A case study is presented where SPOT imagery and GPS ground truth data were integrated using GRASS Software.

In this case study from the Everglades National Park, GPS was used for a research project studying how plant cover in the slash pine forests of the Everglades affects fire management practices. Using GRASS to display and classify a SPOT image, 16 classes of different spectral reflectance were identified. The image was geo-referenced using both USGS quadrangle sheets and a data logging GPS receiver. Excellent results were obtained from each geo-referencing method. The value of GPS became apparent when it was used to navigate to 30 randomly selected ground truth sites. These sites were quickly and easily found and the ground truthing was completed in a very short time. With GPS it was possible to find a site more than 1.5 km from the nearest road through thick underbrush and then travel to four more sites each 0.8 km from the next with complete confidence that the correct site had been located. Navigation in the Everglades pine forest without GPS would have been very difficult.
Introduction
In a project designed to investigate the uses of Global Positioning System (GPS) in Everglades National Park, we conducted a number of tests. These included position data collection, comparison of GPS position data to position data digitized from 7 1/2" topographic maps, and geo-rectification of SPOT imagery using GPS position data. Applications of GPS in Everglades National Park include geo-rectification of remotely sensed imagery, acquisition of position fixes of study sites in remote wilderness, and rapid mapping of long, convoluted plant community and fire boundaries. The liaison between GPS, GRASS, a Geographic Information System, and remotely sensed digital data can enhance the research and resource management activities within a large wilderness, such as the Everglades National Park.

In this study, GPS PATHFINDER™ data recording receivers were used. GPS PATHFINDER receivers allow direct digital data entry, and are easy to use. This ease of use will allow GIS data bases to be created and maintained by users with little knowledge of traditional surveying and mapping techniques.

GPS PATHFINDER is a trademark of Trimble Navigation, Ltd.

Introduction to GPS
When fully deployed, GPS will provide world-wide, 24-hour, 3-dimensional continuous position fixes, accurate to within tens of meters. This system is paid for by the US taxpayer through the US Defense Department. With the current partial coverage and the prospects for 24-hour coverage in a few years, GPS is a viable tool for all concerned with maintaining or creating a GIS data base.

Global Positioning System
A total of 24 GPS satellites are planned, including 3 in-orbit spares, in 6 orbital planes. The orbital period for all the satellites is approximately 12 hours, with an altitude of about 12,000 miles. Each satellite contains a very accurate cesium or rubidium clock which is synchronized to all the others and to the ground control stations. The satellites radiate spread-spectrum coded signals which are received by the user's GPS receiver. Along with timing information, each satellite transmits information allowing computation of the satellite's exact position. The GPS receiver decodes the timing signals from the satellites in view (4 or more), and knowing their location, computes a latitude, longitude, altitude, and time. This is a continuous process and generally the position is updated on a second-by-second basis on the receiver's display.

Current GPS Constellation
The current (January 1990) GPS satellite constellation has eleven satellites. This constellation provides approximately eight hours of 3-D positional fixes per day in North America. Coverage in other parts of the world will vary. The daily observation window appears approximately 4 minutes earlier each day. In the summertime in North America, the current constellation of satellites is visible during daylight hours. In the wintertime, the constellation is visible in North America during the night. As more satellites are launched at the planned 60- to 90-day intervals, these coverage windows will expand until a full 24-hour coverage is available in 2 years.

Older Technologies Being Replaced
The Global Positioning System will eventually replace two current navigation systems, LORAN and TRANSIT. LORAN is a ground based navigation system providing continuous positional fixes accurate to about 500 feet in coastal waters within range of a LORAN chain. The TRANSIT system provides worldwide coverage but only 10 to 30 fixes per day. Sub-meter accuracy is available after 2 days of observation by a stationary user.

Selective Availability and Differential GPS
The autonomous accuracy of positional fixes depends mostly on the actions of the Defense Department and how much they will degrade performance for civilian users. This degradation process is called Selective Availability (SA). When SA is enabled, the position fixes may wander up to 100 meters from the true position. To get a good position fix in the presence of SA, differential GPS is used. With differential GPS, a reference GPS receiver is set up on a known location and any deviations from truth are used to correct the locations computed by another GPS receiver at an unknown location. Systems are available to do this differential processing either in real-time or as a post-processing operation. With differential GPS, accuracy will be on the order of 1 to 5 meters.

If better accuracy is desired, survey grade receivers are available. These receivers are specified to have an accuracy of 1 centimeter + 1 millimeter for each kilometer of baseline. For example, two survey grade receivers separated by 10 km will compute a separation of 10 km ± 2 cm. Survey grade receivers only work in differential mode.

GIS Data Base Benefit of GPS
One of the greatest advantages of GPS for GIS applications is that the data is collected quickly and accurately with a common reference system. With data logging GPS receivers, the coordinates, time, and other attribute information may be collected and then exported to a GIS data base with no manual digitizing operation. Since GPS provides a common reference system, data from GPS sources and sources rectified with GPS will register with each other and with GPS data collected in the field.

GPS Data Recording Receiver
An example of a GPS receiver specifically designed for GIS applications is the Trimble Navigation, Ltd., GPS PATHFINDER system. The GPS PATHFINDER system has a portable, data recording, battery powered GPS receiver for field use and a suite of post-processing programs specifically designed for GIS applications. The most important post-processing function is the ability to export differentially corrected positional data in a desired GIS data format including ARCS/INFO, GRASS, AutoCAD DXF, MOSS, GeoSQL, EPPL7 and ERDAS.
GPS PATHFINDER Operation

While in the field the user can perform a variety of navigation and logging functions. For example, to map the boundary of a field the user would travel to the starting point, turn on the receiver, verify that positional fixes were being generated by observing the display, and turn on the data logger to start collecting a data file. Then all that is necessary is to walk, drive or fly around the boundary to be digitized. At the end of the path the data logger is turned off. Upon returning to the office, the data logger is connected to a PC and the files are extracted and the position data is converted to the desired GIS data import format of your choice. The post-processing software allows reviewing and plotting on 1:24 000, or any other scale, overlays using a HPGL compatible pen plotter.

Using Differential GPS

The GPS data files may be made even more accurate by using differential GPS. In differential GPS, one GPS PATHFINDER receiver is set on a known location and set to record a reference file. This reference file will allow the generation of a time-tagged position correction that can be applied to a remote GPS data file collected within a few hundred miles of the reference location. Using differential GPS, and averaging for three minutes at a stationary point reduces the standard deviation of the error for a point location to less than 2 meters. Laser tracking of a moving vehicle indicates the accuracy of dynamic differential position data will be in the 5 meter range.

Geo-Rectifying a Spot image with the GPS Pathfinder Method

During the nights of 12-14 March 1989, a Trimble Navigation GPS PATHFINDER data logging receiver was used to map features such as hydrologic gauging stations, benchmarks, roads, and plant community boundaries within a south Florida Slash Pine ecosystem approximately 28 000 hectares in area. The GPS position data was converted to a GRASS 3.0 sites format in UTM coordinates, registered to the NAD-27 datum. The four corners of the USGS quad were registered. The residual error for the registration was no more than 3.0 meters. Next, sixteen geo-rectification points on the quad were located and digitized for analysis using a Paired Sample T-test. Digitized point locations from the quad were used to geo-rectify the SPOT image for a comparison to the GPS data. The final scene geo-rectification was done using the GPS position data and the GRASS i.rectify module. The landcover cellfile generated was transferred to a GRASS-GRID Mapset for additional use. The GPS site data was then overlaid on the landcover cell file.

Results and Discussion

The results of the comparison of orthophoto quads (residual mean square error = 0.71) versus GPS (residual mean square error = 0.82) point data for geo-rectification reveals little difference between the two methods. Either error is acceptable for the geo-rectification of SPOT satellite imagery. A direct comparison of GPS data to quad sheet data reveals little difference between the Northing and Eastings. However, a paired Sample t-Test showed a significant difference between the Northing: Mean 7.02 m, std. dev. = 8.70 m, t-value = 3.23, p = 0.006. No significant difference was found between Eastings: Mean = 4.44 m, std. dev. = 18.65, t-value = 0.95, p = 0.366. Although there is a significant difference between Northing and a relatively large standard deviation for the Eastings, these differences are within the error one would expect from digitizing an orthophotocell of scale 1:24 000. At this scale, the width of the crosshairs relative to the ground is about 20 meters.

Ground Truthing Randomly Selected Points

After rectifying the SPOT image, the accuracy of the pixel classification was checked by ground truthing 30 randomly chosen locations. The 30 locations were selected using a random number generator and covered most of the different communities represented in the Long Pine Key study area. The coordinates of the 30 points were converted from NAD-27 UTM coordinates into WGS-84 latitude-longitude positions using the same transformation values originally used to convert from WGS-84 into NAD-27 for the geo-rectification. These positions were used as destination waypoints in the GPS PATHFINDER System. For each destination waypoint, the desired waypoint was navigated to using the real-time navigation display of the GPS set, which shows distance and direction information. A typical point would be 700 meters from the previous point. The pine tree canopy caused a bit of noise in the instantaneous positional fix, however, this was only a problem when we were within about 20 meters of the target. At greater distances, the bearing to the destination was relatively constant. We could tell when we were at the destination point when the bearing started to jump around and the range to the destination was less than 20 feet. At each destination, the vegetation cover was observed and recorded in a field notebook. This information included the density of the canopy, the density of the ground shrubs and the state of the ground cover, since these three different vegetation categories affected the SPOT image classification.

Conclusion

The results of these tests support the use of GPS navigation grade receivers to gather position data for the geo-rectification and ground truthing of satellite imagery. GPS is an operational system that will benefit the users of GIS data bases in collecting accurate geographic coordinates for creating, updating or maintaining their own GIS data bases. Portable data-logging GPS receivers will be especially useful in wilderness areas that have few human cultural features and landforms that appear on satellite imagery but do not appear on orthophotoquad maps.
Data Issues and Remote Sensing

Session F3
Data Structures

A Method of Accessing Large Spatial Databases

Data Structures for GIS

An Efficient Data Structure for Surface Analysis in Vector-based GIS

A Comparison of Partitioned and Seamless Organizations of Spatial Data
A Method of Accessing Large Spatial Databases

Randy Frosh
Genasys II, Inc.
2629 Redwing Road, Suite 330
Fort Collins, CO
80526
USA

Abstract

The explosive growth of GIS technology in the past few years has resulted in production of massive amounts of digital spatial data. When this data is to be used in a coherent regional or state-wide LIS consisting of multiple thematic layers of continuous coverage, accessing specific entities such as an ownership parcel in a timely manner is a common function that demands efficient algorithms. The requirement for efficient spatial searching is recognized as an important GIS topic by The National Center for Geographic Information and Analysis. The National Center has identified development of methods for effectively structuring spatial search algorithms as Objective Two of Research Initiative Six.

In addition to efficient spatial searching, coordinate precision cannot be compromised in these databases nor can project bounds always be set prior to database construction. As different thematic layers require different map resolutions, the same coverage may even require different resolutions due to the density of data. Added to this is the requirement to maintain permanent and temporary views of a partitioned database as a continuous database.

This paper describes methods which use RDBMS and conventional non-DBMS technology for searching and manipulating large spatial databases. The method for sending information to and receiving information from a RDBMS or corporate database is also described.

This method is currently in production use in cities, counties, multi-county regional areas, and statewide areas. The method has been developed by a team of GIS software engineers and fielded in the last revision of a commercial GIS.
Introduction
The requirements for geographic information systems are increasing rapidly. Users are expecting efficient handling of large volumes of data. Additionally they require the system to operate in a continuous or seamless manner. "They simply want to type "PLOT SOILS"" (Reed 1988) and see all soils maps without knowledge of the component map sheets or files.

Generally, spatial database performance depends on the number of disk accesses needed to retrieve data. Disk access is expensive and performance improvements can be achieved by reducing the number of disk accesses through buffering and indexing the storage elements (Egenhofer 1987).

In today's environment of rapid CPU performance increases but slow disk access speed increases, configuring a system whereby the number of disk accesses is minimized leads to a responsive system. The key to efficient accessing of large spatial databases then is to minimize disk access.

Obviously, even a single thematic layer, such as the 3,000 000 land ownership parcels in New South Wales, Australia, cannot be loaded into memory at system initialization time and processed thereafter. While this would minimize disk access to one access for each byte of data, memory prices and the need for concurrent processing of several thematic layers prohibit this approach. A method of indexing and partitioning the database to the users spatial area and theme of interest must be used.

Use A DBMS
Ideally, the solution to efficient access of spatial data lies in currently available DBMS technology. Modern DBMS are fast, offer a variety of indexing methods and handle large databases. However, standard database management systems do not fill this role regarding spatial data. Performance is unacceptable and appropriate data structuring is missing (Egenhofer 1987). Spatial data cannot be properly handled as strings, reals, and integers in a table format. In addition, methods for 'long transactions' are simply not in place in modern DBMS to handle spatial transactions (Dangermond 1989).

Time tests using a DBMS for spatial data on mainframe and supermini-computers can not even match the speeds of a 386-PC. Prototyping of a GIS using a DBMS was quickly abandoned. Typical spatial queries involve comparison of spatial extent with a minimum bounding rectangle. Even this simple operation for a GIS cannot be accomplished efficiently in a generalized DBMS.

In short, at this time, the indexing methods of commercially available DBMS cannot accommodate the needs of spatial data. Even though a DBMS cannot accommodate spatial data as yet another field in a thematic database, they are very adept at handling tabular data structures such as demographic data. This has led to the use of a DBMS in conjunction with a GIS to combine spatial and tabular processing in what has been termed 'Geo-Relational' or 'Hybrid' data models (Morehouse 1985). This is the only practical way for GIS practitioners to handle spatial data unless the goal is to become tabular data experts instead of spatial data experts.

What to Do?
For efficient spatial processing, the solution lies in keeping data that is needed together in close proximity on disk. This can be achieved by computing a spatial index value that combines X- and Y-coordinates as the key indexing field and passing this to a DBMS. The problem is that as map features are added and removed from a map, the DBMS becomes fragmented and data that is needed in one spatial view is in different locations on disk causing loss of efficiency. Even vendors of the most commercially successful DBMS recommend dumping the database to a flat file and reloading on occasion to increase performance.

The file systems of modern operating systems are very adept at keeping blocks of data in readily accessible units. The cost of retrieving a large number of bytes is not much greater than retrieving a small number of bytes once the disk head is in the proper position. If a DBMS cannot efficiently access spatial data, why not use the inherent characteristics of the file system for spatial data instead of all the indexing methods of an expensive DBMS that do not operate effectively on spatial data?

Keep It Simple
Most modern file systems provide for a hierarchical directory structure. The no-cost directory structure provides for efficient access of a multitude of files.

When designing spatial access algorithms, something as generalized as a DBMS is not going to be as efficient for accessing fields of information as static as the structures in a GIS. DBMS are designed for adding fields. A GIS has fixed fields of spatial information and there is seldom a need to add fields to the structures once the system is in production mode.

If commercially available software such as a DBMS cannot provide an efficient method of accessing large spatial databases, custom software must be developed. This software must, however, use proven algorithms that have been researched in the literature and implemented to illustrate effectiveness. To shorten the development cycle and take advantage of other organizations contributions, standard software should also be employed.

The method described in this paper does, for a large part, accomplish the above stated goals. It relies on software such as the UNIX file system that other organizations are enhancing as well as readily available software such as B-tree indexing.

The Implementation
Users are comfortable with the concept of spatial data in layers of information such as soils, hydrology, or parcels. This approach has obvious advantages as pointed out in what has been described as a geo-relational approach as
opposed to a feature-oriented approach (Dueker 1985). Building 'smart software' instead of 'smart databases' has obvious advantages of future flexibility for constructing new approaches to problems and discovering new application needs of spatial data (Dangermond 1989).

Users are also accustomed to separating data into spatial units. Why should data from one disparate place on the earth logically be stored with data from another widely-displaced position?

Using the ideas of thematic coverage and spatial location can be combined with the familiar DBMS concepts of databases or tables. Although loosely defined, the implementation of databases does not prejudice mixing different thematic layers and spatial locations. It simply defines a mechanism whereby users can pigeonhole data. Other groupings can be used instead of groupings by thematic content or spatial location and might include precision, resolution, scale, or classification. Whatever the grouping, from the users point of view the operation is to 'OPEN DATABASE ABC'. The implementation from a software viewpoint is to provide access to a directory of mapped data.

A great deal of access efficiency is achieved very easily using this simple mechanism of separating maps into databases in directories. Only currently open databases are examined allowing the system to ignore large sections of disk.

This approach has at least one distinct advantage, other than efficiency, which can be gained through the methods subsequently described. Access and update control can be assigned to the databases, whether they are spatially, thematically, or otherwise organized. Passwords for read and write access or even operating system access control can be set such that only planners update zoning information and only engineers update water lines, but all users can access parcel boundaries in the cadastral layer.

In the implementation, databases are flexibly defined according to the organization's needs. Read and write passwords are allowed to accommodate access and update. The mechanism also allows for sensitive data such as police information to be under the total control of an individual or department. When read passwords are assigned, data cannot even be accessed unless the password is known. This can also be reinforced by operating system access control for read and write privileges.

Once the database has been established into thematic or spatial groups that can be opened, a method of quickly performing spatial and thematic data searching is required. The method used for locating spatial data incorporates R-tree algorithms (Guttman 1984). The method used for locating thematic data is via B-tree algorithms.

R-trees provide extremely rapid searching of minimum bounding rectangles (MBR). The MBR of each map in the database is loaded to the R-tree index. When a spatial query of the database is conducted, such as 'What data is in my current viewing window?' or 'What data is at this point?', the R-tree index is searched.

B-trees provide extremely rapid searching of thematic attribute information. The primary attribute of a map feature such as the parcel ID of an ownership parcel is loaded to the B-tree index. When a thematic query is conducted, such as 'Retrieve parcel XYZ789', the B-tree is searched.

When a database is constructed, maps can be added to an index map that incorporates R-trees and B-trees. An index map is a map of maps and serve in as a map library. A database typically has one index map but this convention is loosely defined and users are free to construct index maps according to a variety of criteria such as precision, resolution, scale, age, etc. to further dissect a database.

However, there are strict requirements when constructing index maps to provide for a continuous map database. All maps in an index must be in the same map projection and maps must not overlap. Using only these two conditions allows for access of spatial data in a continuous manner as described later in this paper. The map projection must be the same since continuous processing is impossible without use of the same coordinate system. Non-overlapping maps are required for the obvious reason that an overlapping area is not continuous.

Other restrictions that are typical in other GIS are not enforced. No predetermined project extent is established for an index map. The spatial extent of an index map expands as maps are added to it and contracts when maps are removed. Map sheets can be rectangles of any size. No tiling takes place. This allows for smaller spatial extent maps in areas of dense data, such as a city center, as opposed to the periphery of an urban area. Empty tiles do not need to be created for areas that are not yet mapped. No tile size has to be pre-selected, areas of denser data can be in smaller spatial extents so that unnecessary overhead in tile processing is avoided in dense areas. No loss of precision takes place. Each map in an index is stored with its own local offset and precision. In addition, a map is the exact same disk file whether or not it has been added to an index map.

Users can construct a database using a map series typical of paper map coverage. Although users require a modern GIS to operate in a continuous manner, they often request and are more comfortable knowing that the familiar old map sheet XYZMAP or a quad sheet named WOLFRIDGE is still around as a distinct entity. Index maps and the method used do not, however, preclude chopping the spatial coverage into arbitrary rectangles of a fixed size or conglomerating tiles into larger rectangles.

Using R-tree and B-tree algorithms that are proven technology provides a simple but very efficient mechanism for quickly locating a map in a database. Spatial searches are conducted by comparing the MBR of interest to the R-tree. A list of pointers to map partition names is returned. Using
the open database name, a direct path-name is constructed to the spatial data of interest.

In addition, the MBR of interest is first compared to the MBR of the index map before the R-tree of the index is even searched. Speeds available in an R-tree search do not noticeably deteriorate as database size increases since tree re-balancing takes place as rectangles are added and removed from the tree. Response times are approximately one second of clock time for a search of 5000 MBR on a UNIX based workstation such as an HP-370.

All input values are converted to longitude/latitude pairs and MBR in the R-tree are stored as integers. Precision is quite acceptable since .000001 of a degree of precision is stored. All computational operations can then be performed in integer arithmetic which is much faster than floating point. This avoids the need of double precision coordinate storage to maintain map accuracy standards and it also saves disk space by storing thirty-two bit integers instead of sixty-four bit double precision values.

Using a B-tree for thematic search is also very efficient. A primary key is input and compared to the keys in the B-tree index. A list of pointers to map partition names and pointers to features in the map are returned. As a result, a query such as ‘SHOW ME PARCEL XYZ789’ can be accommodated in an extremely rapid manner since a direct pointer to the database/map feature ID is returned. Response times are approximately those of R-trees.

Use of B-trees in an index map is also advantageous in the geo-relational approach where thematic data is stored in a DBMS. After queries are conducted in the DBMS, a common key such as a parcel ID that is the primary key in both the DBMS and the GIS, can be sent from the DBMS to the GIS. The GIS uses the B-tree of the index map and returns not only the map file name but a direct pointer to the feature. This provides for extremely rapid response in thematic/spatial or DBMS/GIS interactions. Similarly, the common key can be sent from the GIS to the DBMS where the DBMS provides its own mechanism for indexing tabular data, performing operations such as relational joins, and forms presentation.

As a side issue, the implementation is not hard-coded to a particular DBMS. Oracle, Ingres, Informix, and HPSQL have all been used. In addition, databases on different machines with different operating systems have been used by transmitting ASCII files of common keys. As a result, existing tabular databases in the DBMS and the computer of the user’s choice can continue to be used. There is no need for a particular DBMS of the GIS vendor’s choice to be forced on the user or to do something as inefficient, limiting, and wasteful as to use one DBMS to access another DBMS.

Another mechanism used to promote efficient access is the use of a sixteen tree spatial index for the frequently searched node table. The nodes in the map are indexed by initially allocating direct references to node records according to their location in a four by four matrix subdivision of the map extent. Each of the sixteen subdivisions matrix elements can directly reference sixteen nodes. When one of the matrix elements needs to index nodes in close proximity that are covered by this individual matrix element, it is further split into sixteen portions and the original matrix element

Accessing Data in an Individual Map
Use of B-trees provides direct references to a feature in a map. This is suitable for efficient access of large databases when the search is a thematic search of the primary feature attribute or common key.

Use of R-trees provides quick reference to the map partitions of interest when the search is a spatial search with an MBR. However, if the map partition contains a large amount of data, such as 20,000 soils polygons, additional methods including fast indexing must be used on the data in the map partition itself to limit disk access.

This is accomplished by several methods. The most obvious and a common method used by popular GIS is segmentation of the map data into tables of information for features, arcs, nodes, primary attributes, etc. This groups data and instills normalization into a dataset. Processing algorithms take advantage of the tables by using the most appropriate table, ignoring the data in the other tables.

Disk buffering in the application program is also used to reduce disk access. The map files are opened with a physical block size equal to the data block size delivered by the disk to the bus. This also results in much less context switching between the application program on the file system when using the UNIX operating system. As a result, relatively large amounts of data (8192 bytes) for each table can be kept memory-resident without the application program becoming unduly large. An example of speeds resulting from segmentation into tables and disk buffering is plotting 1000 topologic street line features per second on an HP-370.

Compressing data so that more map information arrives at every disk access is also used to increase efficiency. One primary example is storing of real world map projection coordinates as either sixteen or thirty-two bit integers. Double precision map projection coordinates are stored internally as local offset values from the center of the map partition. Use of thirty-two bit integers in this manner achieves approximately the same precision as double precision values in half the space. Sixteen bit values provide very acceptable precision in one-fourth the space except for engineering, surveying, or COGO applications. Storing coordinates as integers has the additional advantage of integer arithmetic computational processing. All input double precision search coordinates are first converted to internal integer values allowing computations to take place using integer arithmetic.

Another mechanism used to promote efficient access is the use of a sixteen tree spatial index for the frequently searched node table. The nodes in the map are indexed by initially allocating direct references to node records according to their location in a four by four matrix subdivision of the map extent. Each of the sixteen subdivisions matrix elements can directly reference sixteen nodes. When one of the matrix elements needs to index nodes in close proximity that are covered by this individual matrix element, it is further split into sixteen portions and the original matrix element
points to the smaller sixteen matrix instead of directly to the node. This continues on to a maximum of seven levels where the map resolution is reached. This structure readily accommodates a map with sparse and dense data since each cell is only divided on an as-needed basis. It also results in a maximum of eight disk accesses under the worst mis-disk blocking case in an area of extremely dense data. Speeds achieved for node searches are approximately one clock second in a map of 5000 nodes on an HP 370.

These node table search speeds allow for maintenance of topology during data entry and edit. Gaps and overshots at nodes can be eliminated by tying to existing nodes in the current edit map or ancillary map layers. This also eliminates the need for subsequent cleaning operations to remove the ‘spaghetti’ from the map and does away with nodes ‘creeping’ to some fuzzy position between a set of nodes.

**Permanent and Temporary Views**

Users rarely want to operate on the entire database or even on all map partitions in a thematic layer. As a result, a method of selecting from the databases where specified criteria of features is met must be implemented.

The method operates quickly to allow users to rapidly construct different database views and abandon previous views. For flexibility, the selection method includes not only SQL ‘Select Where’ type conditions but also spatial conditions such as inside of, adjacent to, and in close proximity to. To eliminate re-accessing of data, the separate views can also be combined with the boolean operators AND, OR, NOT and XOR to produce a composite view. Views from selects in the DBMS can be combined with selects in the GIS. Previously constructed views can be input to the select function for further selection.

A simple but inefficient method of storing the views would be to copy the selected spatial data to a new map. This method would obviously increase disk access since the selected data must be written to a disk file and processing would be required to maintain topology. In addition, temporary map views would clutter the user workspace.

Instead of creating temporary maps, the method implements condensed indexing with direct pointers to map features. References to features in multiple databases of multiple map partitions are allowed. The selection criteria is stored with the view to allow for automatic reselection when maps in the view have been updated. A consistent map projection between all the maps potentially in the view is not required since view spatial reference MBR are stored in longitude/latitude pairs and converted to the current session projection ‘on the fly’ when the session projection is different from that of the map partitions in the view. Integer arithmetic is used on views to promote processing efficiency.

Views are treated for all practical purposes as just another map by the analysis and display processing software. The full path to the map partitions are contained in the view with direct feature references. The result is efficient access to spatial data with very minimal additional disk access or processing to read the condensed indexing information in the view.

For flexibility, one or more views can be saved to a new map. In addition, permanent views can be constructed by saving the temporary views. Permanent views can be retained in the user workspace or placed in a database for global access. Temporary views are saved between sessions to eliminate view re-construction.

**Continuous Processing**

A required characteristic of a continuous database is that line and area features which cross map partition boundaries appear to the user as not significantly different from features wholly contained in a single partition. To meet this requirement, tools and procedures are used to ensure that during database construction and editing, the primary attribute and boundary coordinates of a line are the same on both sides of a partition boundary. This capability is provided using the methods previously described and additional processing methods that are not disk- or CPU-intensive.

During construction, searches of the database R-tree quickly establishes which map partitions are in the edit window. As data is entered, it is a simple matter to compare the MBR of the incoming data to the MBR of each partition and drop the data into the appropriate partition. The incoming data and the MBR of the partitions are memory-resident and as a result, the scheme requires minimal overhead and operates efficiently.

Lines which cross a partition boundary are divided at the boundary and flagged as ones which cross. During edit, query, or select operations, the composite line can quickly be accessed by using the highly efficient node index. The coordinates of the boundary node identify the adjacent partition by a simple MBR search. The node index of the adjacent partition is then searched and the line attached to the node is returned. This continues across partitions until a line is encountered which does not cross partition boundaries or has a different primary attribute. For area features, since they are formed from edges used in a clockwise direction, the method simply ‘turns right’ at the adjacent partition boundary node.

The degree of efficiency in these operations results in very little differences in response times between accessing features crossing partition boundaries and ones wholly contained in a single partition. In addition, the map sheets can be formed directly without making a ‘mongo map’ and tiling into a database.

For data coming from a non-topologic system or a GIS without sufficient edge matching, zippering functions are provided to identify boundary node coordinate mismatches and primary attribute coordinate mismatches across partition boundaries. The effect of the functions is to run a zipper around the map borders and determine where the teeth(nodes) do not match. Zippering uses the node index which result in its high efficiency.
Conclusion

Efficient access to a continuous spatial data is provided by the methods described. Limiting disk access through partitioning, indexing, and compression, along with computational prudence results in the rapid response times. The method uses straightforward algorithms and does not require the user to bear the cost of purchasing a license to use a DBMS to make the GIS operate.

Although current DBMS technology cannot efficiently process spatial data, it is expected that this will not be the case in the near future. Already some DBMS are beginning to show promise. Perhaps some of the indexing algorithms presented here could be implemented in a DBMS for fields of information with spatial characteristics.

References


GIS'90 Symposium

Vancouver, British Columbia
March 1990

Data Structures for GIS

Karen Westwood
GeoVision Corporation
Ottawa, Ontario
Canada

Abstract
How data structures are implemented and integrated has critical ramifications on system design, capability, efficiency, expandability, and longevity. Structures which consider the nature of the data they are storing are the key to any successful system, particularly with spatially based information, due to the complex nature of geographical relationships. This paper reviews hierarchical, network, and relational data structures as well as raster and vector data models. It also provides examples and evaluates the integration of these data structures and models into commercial GISs.
An Efficient Data Structure for Surface Analysis in Vector-based GIS

Yue Hong Chou
Assistant Professor
Department of Earth Sciences
University of California
Riverside, California
92521
USA

Abstract
Surface analyses based on digital topographic data, using a geographic information system (GIS), often involve the processing of large data sets with complicated operational procedures. The design of an efficient data structure improves the effectiveness of a GIS in handling surface information. In general, vector data structures are more efficient than raster structures concerning memory space and processing speed. Three types of vector structures are available for organizing surface information: a point coverage with irregularly spaced data points, a polygon coverage of triangulated irregular networks, and a line coverage of digitized contours. Comparisons among them indicate that digitized contours provide the most efficient data structure. Two relational files are used to construct an efficient structure of digitized contours, a Contour Attribute Table containing topographical and topological information and a Point Location File containing locational data of contours. A surface model for detecting slope lines from digital terrain models is built on this data structure and applied to the topographic database of Idyllwild, California. The data structure enables surface information to be processed with minimally required topographic and topological data elements and thus improves the efficiency of complicated computations in a variety of surface analyses.
Introduction

Three fundamental elements of topographic data are required in surface analyses, the X and Y coordinates of each data point and its elevation conventionally denoted as Z. Most digital topographic data are organized in a raster environment where data points are regularly spaced in X and Y dimensions; e.g., surface elevation is coded at a constant distance of thirty meters in the Digital Elevation Models (DEM) of U.S. Geological Survey (1987). Earlier Geographic Information Systems (GIS) are developed accordingly and organize surface data in raster structures.

Raster data structures are easy to implement and ready for grid-cell processing. Regarding memory space, an obvious advantage of raster structures is that, as data are consistently arranged, only the Z values need to be coded as long as X and Y values are expressed explicitly by a simple function of regularity. However, due to the vast amount of redundant data, empirical topographic databases organized in raster structures tend to require much more memory space than what they save on uncoded X and Y values. Particularly, for large flat surfaces or areas of relatively constant slopes, raster data files usually contain a great deal of data points that are either of identical values or are varying constantly and systematically.

A major criterion for a surface analysis to be efficient is that topological characteristics of areas beyond the boundary of grid cells must be readily available. Simple raster structures are inadequate for surface analyses because, as topological information is incorporated in such structures, the size of the database increases enormously. In some cases, the volume may become so large that even basic procedures of data processing take a great deal of CPU time. Storage-efficient raster structures are available, e.g., the quadtree structure (Mark et al. 1989), yet the saving in memory space is usually obtained at the cost of increased complexity in coding topological information. Consequently, it is very inefficient to conduct any surface analysis which requires frequent processing of topological characteristics based on the modified raster structures.

In general, vector structures are more efficient in terms of data storage primarily because spatial data are organized with variable density. For instance, a large flat surface may be coded by a single data point whereas a constant slope may be accurately represented by a few data points. Surface analyses can be performed much more efficiently in vector-based GIS as a result of minimized data redundancy and improved processing speed. In this paper, major vector structures of topographic data are compared in terms of storage space and processing speed. A data structure designed for efficient surface analyses is presented.

Vector Structures of Topographic Data

In vector-based GIS, topographic data can be organized into either a point coverage, a line coverage of digitized contours, or a polygon coverage of irregularly triangulated networks (TIN) (Peucker and Chrisman 1975, Estes 1985, Dangermond 1988).

A point coverage of topographic data contains the X, Y, (location) and Z (elevation) values of irregularly spaced data points. The coverage may be rearranged through interpolation to generate a lattice coverage of regularly spaced data points. For surface analyses, the data structure and processing procedure of the lattice coverage are similar to those of a raster database thus the lattice coverage is not considered here. Figure 1 shows the coverage of topographic data points of the Palmview Peak 7.5-minute quadrangle in southern California. This coverage is clipped out from the database of Idyllwild 15-minute quadrangle. The original Idyllwild database is created by converting the DEM into a lattice coverage and processed through the selecting procedure of Very Important Points (VIP) to reduce data redundancy (Environmental Systems Research Institute 1987, Chen and Guevara 1987). There are 719,352 data points in the original DEM. The processed point coverage contains 1,156 data points.

Figure 1: The point coverage of Palmview Peak, northeast Idyllwild, California. Each dot denotes the location of one data point coded with elevation.

Figure 2 shows the line coverage of digitized contours generated from the point coverage. Data files of the contour coverage contain line segments of contours, their elevations, and the X and Y coordinates of the points in the segments. A code of topological information is assigned to each point to differentiate starting and end nodes from intermediate vertices.
Figure 3 shows the polygon coverage of triangulated irregular networks (TIN) generated from the point coverage. The TIN coverage is a special case of polygon coverages for all polygons are degenerated into triangular facets. Topographic data are coded on the basis of the triangular facets.

Comparisons among Vector Structures
Storage space is the first element to be considered in assessing the efficiency of a data structure for surface analyses. Table 1 shows the volumes of data files organized in vector structures.

Table 1: Data Volumes of Vector Structures

<table>
<thead>
<tr>
<th>Coverage</th>
<th>Data Volumes (k bytes)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Palmview (7.5)</td>
</tr>
<tr>
<td>Lattice (DEM)</td>
<td>720</td>
</tr>
<tr>
<td>Point</td>
<td>68</td>
</tr>
<tr>
<td>TIN</td>
<td>606</td>
</tr>
<tr>
<td>Contour (100m)</td>
<td>124</td>
</tr>
<tr>
<td>CAT-PLF (100m)</td>
<td>83</td>
</tr>
</tbody>
</table>

It should be mentioned that comparing the size of the lattice coverage with that of any other coverage is inappropriate because the lattice coverage is converted from the DEM without generalizing while other coverages are based on the data set processed through the VIP procedure. The lattice coverage contains the same number of data points as the unprocessed DEM and is listed only for demonstrating the size of DEM. CAT-PLF represents the proposed data structure derived from the line coverage of 100-meter contours, which will be discussed later.

As far as storage space is concerned, the point coverage is most efficient among vector structures due to the fact that topological information is not incorporated. However, the lack of topological information also makes the point coverage least useful for any surface analysis. This is the simplest data structure suitable only for unprocessed raw data. Furthermore, the map of a point coverage can be hardly interpreted for any purpose. In order to conduct any surface analysis, the point coverage always has to be converted into either a lattice coverage, a contour coverage, or a TIN coverage. Clearly, the advantage of minimal storage space of the point coverage is useful only when the topographic data are to be archived. Since the cost of disk space is relatively low, it is more efficient to organize topographic data in a structure readily available for surface analyses.

Digitized contours and TIN are the commonly accepted means for organizing topographic data (Peucker et al. 1978, Carter 1988). The TIN coverage has certain important features. First, a triangular facet is the smallest indivisible surface with unique and unambiguous surface properties. Second, the number of triangular facets in a coverage is a function of data points. If \( N \) denotes the number of data points, then the number of unique triangular facets is between \( N-2 \) and \( 2(N-3)+1 \). This property enables the analyst to adjust the size of a TIN coverage by choosing an adequate number of data points. Another useful feature of
the TIN structure is that surface characteristics of individual triangular facets may be derived mathematically. For instance, the surface area, planimetric area, slope gradient, and slope aspect (orientation) of every triangular facet are available in the TIN coverage. However, these surface characteristics are of little use for surface analyses because their meanings are limited to the area within each triangular facet.

Certain disadvantages of the TIN structure hamper its efficiency in surface analyses. First, the TIN structure is sensitive to edge effects. Since the formation of triangulated networks is bounded by a convex hull, an additional data point near the edge of the coverage may alter the structure of a large portion of the TIN coverage.

Second, the file structure of the TIN coverage is cumbersome. The topographic and topological data may be organized either in a number of relational TIN files or in two attribute tables of a polygon coverage. For instance, in ARC/INFO, a commonly used vector-based GIS, six relational files are constructed for each set of TIN, including the environment file (ENV), the file of the convex hull (HUL), the file of edges (EDG), the file of vertices (NOD), the files of X and Y coordinates (NXY), and the file of elevation data (NZ). Surface analyses often require the processing of surface information beyond the boundary of triangular facets. The cumbersome structure of TIN files creates stumbling blocks for basic neighborhood operations.

Answers to queries about the spatial relationship of neighboring triangular facets usually are reached after a number of tedious logical operations. For instance, simple questions such as whether the path along a slope reaches an edge or a vertex or which adjacent triangle should be considered in extending the path may cost a great deal of CPU time in processing through TIN files.

Another disadvantage of the TIN structure is that triangular facets are poor representations of the Earth's surface. It is difficult to identify any spatial pattern from a map of triangulated networks. The output of a TIN coverage must be converted into more understandable forms such as contours or perspective diagrams.

Digitized contours not only provide an efficient structure of topographic data regarding storage requirement (Douglas 1986), they are also ready for graphic presentation and surface analyses. In a vector GIS, a line coverage of digitized contours contains topographic and topological data in an Arc Attribute Table (AAT) which is not designed specifically for surface analyses. A data structure derived from the contour coverage for improving the efficiency of surface analyses is presented below.

The CAT-PLF Data Structure
Topographic data can be organized in two relational files, the Contour Attribute Table (CAT) and the Point Location File (PLF). The PLF is a sequential list of locations of all the points that construct the contours, coded in X and Y coordinates. Table 2 shows a partial list of the PLF of the Idylwild contour coverage. This table does not include the addresses of data points, which may be added in order to enhance processing efficiency with binary search.

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>543143</td>
<td>3734040</td>
<td>542876</td>
<td>3734085</td>
</tr>
<tr>
<td>524396</td>
<td>3734165</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>524499</td>
<td>3707473</td>
<td>524044</td>
<td>3706968</td>
</tr>
</tbody>
</table>

The CAT contains the topographic and topological data of contours. Table 3 shows a partial list of the CAT of the Idylwild coverage. Each record is an arc of a contour labeled by ID. Five fixed-length items are coded: elevation (ELEV), address of the starting node (START), the number of data points (PNT), the number of neighboring contour segments of lower elevation (LC), and the number of neighboring contour segments of higher elevation (HC). Two items of variable length are also coded, the list of neighboring lower contours (LCS) and the list of neighboring higher contours (HCS), in both lists contours are referenced by ID codes. Since the lists of higher and lower neighboring contours are of variable length, they may be saved in separate files for convenience.

<table>
<thead>
<tr>
<th>ID</th>
<th>ELEV</th>
<th>START</th>
<th>PNT</th>
<th>LC</th>
<th>HC</th>
<th>LCS</th>
<th>HCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>200</td>
<td>1</td>
<td>22</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>1800</td>
<td>23</td>
<td>47</td>
<td>2</td>
<td>0</td>
<td>6,30</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>800</td>
<td>71</td>
<td>49</td>
<td>1</td>
<td>2</td>
<td>17</td>
<td>4,12</td>
</tr>
</tbody>
</table>

START is the relational key between the CAT and the PLF. The locations of vertices in each contour segment are sequentially arranged in the PLF and referenced by the address code of START in CAT. LCS and HCS are the critical elements of the CAT-PLF structure, providing the spatial relationships among segments of neighboring contours. The lists of neighboring contours are important for efficient surface analyses because they minimize the choice set in each local search and thus reduce CPU time in processing topographic data. The extraction of surface information is performed only on necessary records instead of the entire database. The efficiency of surface analyses is further improved by the minimized data redundancy in the CAT-PLF structure. Table 1 shows that this structure is more efficient in data storage than the original contour coverage.

Surface Analyses using the CAT-PLF Structure
Complicated procedures in surface analyses can be performed efficiently using the CAT-PLF structure. Based on this structure, Chou (1999) proposes a method for vector-based GIS to detect the path of steepest descent from any exogenously given location in a surface. The method is developed for detecting the most likely routes of landslide debris along slopes, which is useful for identifying potential sites of deposit and delineating zones demanding...
slope protection. The problem is complicated because it involves the processing of both topographic and topological data of the coverage. The Idyllwild database is used for testing the method. A 200-meter contour coverage is generalized from the 100-meter contour coverage and then converted into the CAT-PLF files through a FORTRAN program. Chou's method correctly identifies the slope line starting from any arbitrarily given location in an average of 1.6 seconds CPU time on a VAX8820 computer. The efficiency of this procedure relies heavily on the high processing speed contributed by the CAT-PLF data structure.

Other procedures of surface analyses can also be performed efficiently using the CAT-PLF structure. For instance, to generate a perspective diagram (3-dimensional graph) from a contour coverage, one may impose a fishnet grid on the contours coded in the CAT-PLF structure. The topographic profile can be constructed from the intersections between contours and each grid line. Intersections can be identified efficiently because the lists of neighboring contours reduce the processing time in searching for next intersections. Each time an intersection is identified, the next intersection is always on the neighboring contour. Another example is automated generation of intermediate contours. Since intermediate contours exist only between neighboring contours, the lists of neighboring contours minimize the choice sets for data processing thereby saving a great deal of CPU time.

Conclusion
The current GIS technology enables a wide variety of spatial analyses to be conducted efficiently. However, different data structures are suitable for different types of analyses; for instance, the same overlay analysis that involves complicated logical and mathematical operations in a vector data structure may be more efficiently performed in basic raster structures. A well designed data structure not only improves the efficiency of a spatial analysis, but also opens the door to more sophisticated analytical procedures.

The major concern in surface analyses is the distribution and the variation of elevation over the X-Y space. As both storage space and processing speed are considered, digitized contours provide the most efficient data structure. The CAT-PLF structure is based on digitized contours and designed specifically for complicated procedures in surface analyses.

Since different data structures are suitable for different spatial analyses while an efficient GIS database can not be constructed in all structures, an important issue for GIS developers is the capability of data conversion among different structures. All the data conversions in this study are automated. The DEM are converted into a contour coverage by ARC/INFO procedures. The Arc Attribute Table (AAT) and the Polygon Attribute Table (PAT) of the contour coverage are then converted into the CAT and PLF files by a conversion program. The next generation GIS must be versatile for different types of spatial analyses such as overlay analyses, statistical analyses, spatial modeling, and surface analyses. Even if standardized data structures are adopted by major GIS in the future, it is still necessary to maintain the data-conversion capability for different spatial analyses to be conducted efficiently.

Acknowledgement
This research is sponsored by US Forest Service under Grant PSW-90-0004CA. The author is grateful to Lucy A. Salazar of Riverside Forest Fire Laboratory, US Forest Service for providing the Idyllwild database. The methodology is built upon a previous project sponsored by US Forest Service under Grant PSW-89-0001CA. Robert R. Ziemen of US Forest Service provided important assistance.

References


A Comparison of Partitioned and Seamless Organizations of Spatial Data

Stan Cabay, Luca Vanzella
Department of Computing Science
University of Alberta
Edmonton, Alberta
T6G 2H1
Canada

Abstract
For two-dimensional vector-encoded spatial data, consider the partitioning of the area of interest into a set of adjacent tiles. If the spatial data itself is split according to the tile boundaries and the tiles are organized in some manner, then the organization is said to be partitioned. If the spatial data remains unaffected by the introduction of the tiles and the tiles serve primarily as an index to the vector-encoded data, then the organization is said to be seamless. This paper compares the costs associated with partitioned and seamless organizations for performing some common operations on spatial data sets, with particular emphasis on very large thematic data sets. To put comparisons on a concrete foundation, a particular but representative realization of a partitioned and a seamless organization is described.
1. Introduction
For vector-encoded spatial data, the perception of some researchers and vendors of geographic information systems is that seamless models are unquestionably the desired objective [7]. The main contention is that the world is not partitioned into discrete rectangular sections corresponding to mapsheets, and there is no reason why we should not model the world as it really is continuous [1]. From a practical point of view, the argument is that partitioning by mapsheets introduces storage overhead and makes certain operations such as tracing and connectivity analysis tedious if not impossible [13].

We acknowledge that the seamless database is clean and simple, and that for certain spatial operations it may be preferred. However, in the following, we attempt to convince the reader that the controversy over the seamless versus the partitioned approaches is very much alive. Indeed, arguments are given which would indicate that for large thematic databases there are very good reasons for partitioning by mapsheets, and that this is possible without significantly affecting either storage requirements or costs of operations.

The strength of the partitioned spatial database comes from its ability to adopt the divide-and-conquer approach to retrieval and processing [18,19,21,23] The partitioning of the study area into mapsheets (tiles) is the division aspect. This aspect introduces artificial information into the database, namely, the information along the boundaries of the tiles (the seams of the partitioned database). The conquer aspect is achieved by storing sufficient information in each tile so that spatial operations can be performed on a tilewise (tile by tile) basis. Each tile is independently retrieved and processed. In addition, within each tile sufficient topological information can be retained about neighbouring tiles to permit the effective execution of topological operations such as tracing and connectivity analysis. The complexity introduced into the database by the additional information at the seams is the price that is paid for any improved performance obtained from the conquer aspect. In the following, we argue that for large thematic data sets this price need not be high.

2. A Topological Vector Organization
The underlying assumption in the following discussion is that the spatial data is encoded in a vector format. The merits of such a format as opposed to a raster one is an ongoing controversy and are discussed elsewhere [2,16]. The scope of spatial data considered in this paper is restricted to data describing collections of regions (objects) in two-dimensional space. We are primarily interested in Geographic Information Systems (GIS) dealing with collections that totally partition an area of interest into a set of mutually-exclusive regions. Typical of these applications is the handling of thematic maps that describe phenomena areally distributed on the surface of the Earth. Examples of such thematic data applications are land use classifications, forest covers, and soil types. Of primary interest are very large thematic data sets which consequently must reside on secondary storage.

There are a number of choices for the format of vector encoded spatial data [2,16,22]. However, in order that certain operations such as tracing be performed in an effective manner, it is paramount [13] that the format include topological information, that is, spatial and attribute information about neighbourhoods. We briefly describe one such set of formats; this set is typical of the formats adopted by well-known topological GIS's such as POLYVRT[15], GEOVISION[13] and especially TIGER[12]. These formats are used as the underlying vector data structure for describing in the next section the seamless and partitioned data models.

A polygon network corresponding to a theme (or coverage) can be represented using three entity types; polygons, chains and nodes. A chain is a sequence of edges defining a boundary of a polygon (usually separating two polygons). It is typically represented by an ordered sequence of points corresponding to the beginning and end points of the edges that make up the chain. The nodes in a network are the first and last points of all chains in the network. The nodes are the meeting points of three or more polygons (i.e., the meeting points of three or more chains).

The central elements of a topological vector data model are the chains, which are usually stored randomly in a chain file. Each record in the chain file gives the start and end nodes and the left/right polygons of a particular chain. In addition, each chain record points to two other chains having a node in common with the start and end nodes. This technique gives the ability to determine all the chains intersecting a particular node and enables the threading[12] of chains around a given polygon. The coordinates of the points composing a chain are not explicitly stored with the chain record; instead a pointer is made to a location in a separate file, the point file, which lists these coordinates. Such a strategy allows certain operations, such as connectivity analysis, to be performed without requiring the costly retrieval of the coordinates of chains.

To provide access to the polygons in the network, a polygon file can be used. Each record in this file corresponds to a distinct polygon in the network. Each record stores a pointer to a chain in the chain file which has the associated polygon on its left or right side. To accommodate polygons with holes, further organization of the polygons in the polygon file is necessary. Weiler[24], for example, uses two additional pointers in each record of the polygon file. One pointer provides a link to a polygon which coexists (neither lies within the area of the other), while the other refers to a polygon which is entirely contained within (a hole).

The structure so far described does not yet provide easy access to the spatial extent of any particular attribute of a theme (e.g., in a forestry coverage, the polygons whose attribute might be 'spruce'). This can be enhanced by placing appropriate attribute information in the records of the polygon file. The polygons having the same attribute can then be linked by means of pointers, and a separate file can be set up, which, for a particular attribute, points to the first polygon with that attribute.
Alternatively, the separate file can, for each attribute, list all the polygons in the polygon file which possess this attribute.

Many spatial operations can be performed effectively using the organization outlined above. This is true in particular for operations which in [21] are classified as attribute-based (given an attribute, what are its locations?). Thus, for example, given a polygon or an attribute of a coverage (which is equivalent to a list of polygons), it is necessary only to thread through the chain file to determine the chains composing the boundaries of that polygon or attribute. The boundaries are then obtained by retrieving from the point file only those coordinates describing the relevant chains; no extraneous information is retrieved. Also, because of the topological content within the chain file, the topological operations of tracing and connectivity analysis can be performed in an efficient manner.

Location-based operations [21] (given a location, what are its attributes?) are more costly. Included in this class of operations are point inclusion and windowing. To improve the situation, the TIGER system, for example, includes another file which we here call the node file. Each record in the node file corresponds to one node in the polygon network. The record gives the coordinates of the node and a pointer to one chain that emanates from that node. To assist in location-based operations, a spatial ordering by Peano keys [5] can be imposed on the records of the node file. Binary search methods can then be used to help identify nodes relevant to a location-based operation. For the point inclusion operation, using the Peano key of a given point allows rapid determination of a network node nearest to the point (i.e., nearest in Peano order). The set of chains emanating from the nearest node are obtained from the threaded chains in the chain file. This set of chains divides the space around the nearest node into neighborhoods, one of which may be the polygon containing the point. If none of the neighborhoods contain the point, then this method has at least identified a reasonable location within the polygon network from which the search for the enclosing polygon may begin. For the windowing operation, this node file organization provides similar improvements in performance. By extracting from the node file the set of nodes contained within the window (note that this set may be empty), the search for candidate polygons intersecting the window is localized to the neighborhoods of the nodes within this set.

However, even with these refinements the spatial data model described above, which is entirely vectorial, is completely inadequate for the thematic overlay operation. This operation is both attribute-based and location-based and is called a hybrid operation in [21]. Thematic overlays involve the complete traversal of the polygon networks representing two or more themes. Restricted by the lack of information on the location of polygons, for large thematic data sets, the overlay problem typically involves multiple retrievals of polygons and their boundaries.

The quest to improve the performance of the polygon overlay and the location-based operations has been the primary motivation for introducing additional structures on top of the underlying vector one. One such structure is the subject of the next section and the theme of this paper; another which is worthy of note is the R-tree [8]. The R-tree requires that the polygons in a polygon network be tightly enclosed by rectangles; these provide estimates for the locations of the polygons in space. The coordinates of the rectangles are concatenated to the records of the polygon file. The polygon file can now be spatially ordered and hierarchically organized by the bounding rectangles (the R-tree), providing improved access to those polygons that intersect a specified location.

3. The Seamless and Partitioned Data Organizations
For vector-encoded spatial databases, a method of improving the efficiency of the polygon overlay and location-based operations is to introduce a tessellation of the underlying space into a exhaustive collection of mutually exclusive cells. Each cell either references those polygons that are contained within it or intersects it (the seamless approach), or each cell contains that portion of the polygon network that intersects it (the partitioned approach).

In practice, it is usually desirable to select cells to be rectangular (the cells are then also called tiles, mapshets or frames) and to allow the cell sizes to vary according to some hierarchical subdivision principle. For rectangular cells, the subdivision principle allows the cells to be organized into a quadtree. Many benefits accrue from a quadtree organization, the most significant being the ability to use direct addressing techniques to access cells (or, the buckets associated with them)[10,14,20]. There are various strategies for selecting tile sizes depending on such criterion as the local density of data [6,21]. In the following, no assumptions on the tile sizes are made, only that they are identical for both the seamless and the partitioned approaches. This allows comparative arguments in the next section to be made more easily.

The splitting of a polygon network by a tessellation grid can be achieved in a number of ways. Each time a tile boundary intersects a chain, the chain is split into two new shorter chains. Each such occurrence introduces one additional chain to the chain file, one additional node to the node file, and two additional points to the point file (one end point each for the two new chains which meet at the new node). For a polygon network, this splitting process can significantly increase the size of the chain and node files and mildly increase the size of the point file. However, for typical thematic data sets it is not unreasonable to assume that the storage requirements for the point file overwhelmingly dominate that of the chain and node files. It can therefore be argued that the partitioning process has only a moderate effect on the storage required for these three files.

The partitioning of a network also splits polygons into a number of smaller ones (see Figure 1). If processing is to be possible on a tilewise basis, it is necessary to organize these polygon portions within each file. For the network configured in Figure 1, a Weiler-type organization of the polygon portions is illustrated in Figure 2. In Figure 2, a
Note that chains introduced by the tile boundaries, which are necessary for defining the extents of polygon portions, are excluded from the chain file. Boundary chains (seams) can easily be computed whenever they are required (assuming, not too unrealistically, that the coordinates of the tile boundaries are available). For example, in Figure 1, it is easily determined that the polygon portion of P4 which has chain 11 as part of boundary can be completely defined by joining one of its endpoints to the bottom right corner of the tile and then joining this to the other endpoint of chain 11. In Table 1, chains with asterisks are those in neighbouring mapsheets which are connected to chains in the present tile.

Also note that the location of a node introduced at a tile boundary is required only for operations which also process the coordinates of one or more chains meeting at that node. In this case, the coordinates of such a node are already available from the chain being processed. Thus, after all, there really is no need to store in the node file the nodes introduced at tile boundaries; that is, partitioning need not affect the node file (if a node file is used).

Partitioning, however, does have an effect on the polygon file. Each record in the file, which corresponds to a polygon, now points to a tile which intersects the boundary of the polygon, rather than to a chain composing its boundary.

The crucial observation can now be made. In the partitioned approach, it is possible to arrange that each record in any of the polygon/cell, chain, node and point files is associated with precisely one tile. The tiles are quite independent (the one exception is that chains in one tile contain pointers to connecting chains in neighbouring tiles). This provides the ability to do tilewise ordering of the records (by Peano keys, say) of all four files. A preferred organization is to assign a bucket to each tile and use direct addressing techniques to retrieve in one input/output operation (I/O) all the records for that tile; this arrangement is assumed in the next section. However, in some environments, it may be better to use a separate bucket organization for each file.

4. Comparison of Costs of Operations
The tessellation component (i.e., the polygon/cell file) of the seamless approach is clean and simple. It serves simply as an index to attribute (e.g., polygon) information intersecting each of the tiles in the tessellation. The underlying polygon network remains intact. On the other hand, in the partitioned approach, some complexities in the data structure are introduced by the splitting of the polygon network at the tile boundaries. In addition, there is a significant increase in the storage required by some of the files in the partitioned approach, although it is argued in Section 3 that, overall, this increase for typical thematic data sets is modest. The question remains: using the partitioned model, is there adequate compensation for this increase in complexity and storage requirements?

To this end, the two approaches are compared in terms of the costs of some common operations performed on thematic data. The measure of cost is the number of I/Os required to complete an operation (the processing costs
associated with each of the given operations are assumed to be less important).

For an attribute-based operation, consider the retrieval of the boundary points of a specific polygon. Using the seamless approach, the polygon file is first accessed to obtain one chain lying on the boundary of the polygon. Assume that this can be accomplished in one I/O (using, for example, direct addressing techniques). The remaining chains along the boundary of the polygon are obtained by traversing the appropriate chains in the chain file. Since the chain file cannot be spatially ordered in a meaningful way, a separate I/O is required for each chain so retrieved, in the worst case. (Some improvement may be possible here if, say, the chains are spatially ordered by the coordinates of one of their endpoints.) In addition, a separate I/O is required to obtain from the point file the coordinates of the points composing each chain. Thus, the retrieval of the boundary points of a polygon with c chains using the seamless model normally requires 2c+1 I/Os in the worst case. On the other hand, using the partitioned approach, again one I/O is required to access first the polygon file to determine one tile which intersects the polygon. By traversing the chains within this tile, other tiles intersecting the given polygon are obtained and traversed. Therefore, if the polygon intersects t tiles, its boundary points using the partitioned approach can always be obtained in t+1 I/Os. Thus, the costs of two approaches for polygon retrieval are not comparable; 2c+1 may be larger or smaller than t+1.

Summarizing, in the seamless model, the cost of the operation depends on the number of chains involved in the seamless approach, the set of candidate polygons which intersects the given tile are first obtained by accessing the polygon/cell file. The candidate polygons are then sequentially processed, each requiring the separate retrieval of all the chains and points composing its boundaries. In the worst case, all of the candidate chains must be retrieved. Similar advantages of the partitioned approach are realized for the windowing operation. The tiles intersecting the window are determined (often simply by computation), retrieved and individually processed. For every tile retrieved, some information is always relevant to the window. Using the seamless approach, all the polygons (their chains and points) which intersect the window must be retrieved and processed. However, in this case, not every polygon (chain) retrieved is relevant to the window.

The greatest advantage of the partitioned approach is realized for the thematic overlay operation. The corresponding tiles of two or more coverages are individually retrieved and overlaid (the power of the divide-and-conquer capability is fully realized here). If the sizes of the tiles from separate themes do not correspond, the larger tile may require splitting (this assumes the same decomposition principle is used for both themes). Using the seamless approach, an overlay is performed of all polygons which intersect a particular tile. This requires numerous I/Os in comparison to the two I/Os (assuming two themes) required by the partitioned approach. Since some of the polygons in this tile have influence in neighbouring tiles not yet processed, it is desirable to retain these in memory. For very large thematic overlays, however, this can lead to memory overflows.

<table>
<thead>
<tr>
<th>Chain</th>
<th>Start Node</th>
<th>Start Pointer</th>
<th>End Node</th>
<th>End Pointer</th>
<th>Left Attribute</th>
<th>Right Attribute</th>
<th>Coordinates Pointer</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>A</td>
<td>1*</td>
<td>B</td>
<td>3*</td>
<td>P1</td>
<td>P3</td>
<td>→</td>
</tr>
<tr>
<td>4</td>
<td>C</td>
<td>3*</td>
<td>D</td>
<td>17</td>
<td>P1</td>
<td>P3</td>
<td>→</td>
</tr>
<tr>
<td>5</td>
<td>D</td>
<td>4</td>
<td>E</td>
<td>6*</td>
<td>P1</td>
<td>P4</td>
<td>→</td>
</tr>
<tr>
<td>9</td>
<td>H</td>
<td>10*</td>
<td>G</td>
<td>8*</td>
<td>P4</td>
<td>P5</td>
<td>→</td>
</tr>
<tr>
<td>11</td>
<td>J</td>
<td>12*</td>
<td>I</td>
<td>10*</td>
<td>P4</td>
<td>P5</td>
<td>→</td>
</tr>
<tr>
<td>17</td>
<td>D</td>
<td>5</td>
<td>M</td>
<td>16*</td>
<td>P4</td>
<td>P3</td>
<td>→</td>
</tr>
<tr>
<td>18</td>
<td>N</td>
<td>18</td>
<td>N</td>
<td>18</td>
<td>P2</td>
<td>P3</td>
<td>→</td>
</tr>
</tbody>
</table>

Table 1

Chain File

operation, whereas in the partitioned model the cost depends on the number of tiles involved. This is a general observation applicable to all operations performed on the polygon network.

Thus, a similar observation holds for a topological operation such as tracing. If the thread that is traced is composed of c chains and intersects t mapsheets, then the cost using the seamless model is 2c I/Os in the worst case as opposed to t I/Os using the partitioned one.

Significant advantages of the partitioned approach are realized for location-based operations such as point inclusion. The tile containing the given point is first determined (computed). The data for this tile is retrieved and processed; only one I/O is required. On the other hand, using the

Consequently, multiple retrievals of polygons may be required, adding further to the cost of an operation that is already very expensive.

5. Conclusions

A frequent objection to the partitioned spatial data model is that it may require user interaction with tiles. Although this is a feature provided by some commercial systems, it is possible and probably desirable to have the seams introduced at the tile boundaries appear transparent to the user. The main purpose for the partitioning is to provide a physical organization of the data, which gives a performance level that is basically independent of data volume. That is,

1 The arguments and conclusions of this paper are made in the absence of the Proceedings of the Symposium on the Design and Implementation of Large Spatial Data Bases, NCGIA, Santa Barbara, California, July 1989.
the partitions are meant to supply a physical view of the
data, and should not interfere with the logical view. The
various components of the polygon network (e.g., the node
file) can still be placed in the context of a relational data-
base [3, 11, 17].

Another objection to the partitioned approach is that it is
undesirable to split objects (polygons) at tile boundaries.
However, there is no difficulty in reconstructing a polygon.
Just as for an unpartitioned vector model, it is necessary
simply to traverse the boundary chains of the polygon.
Whether or not this increases or decreases costs depends
on the make-up of the polygon and of the tessellation (see
Section 4).

Still another objection to the partitioned approach refers to
the additional costs incurred during the splitting and merg-
ing of tiles. However, these costs are not prohibitive. The
splitting of a tile can be viewed simply as an overlay
operation, which should not be too costly since processing
is localized to the tile being split. The merging of tiles can be
accomplished by traversing the polygons within the tile,
joining the required polygon portions and chain segments
during the process. Note that for processing purposes it is
not necessary to merge the chains segments at all; indeed,
there are some advantages to leaving the chains in their
segmented form.

There are other potential advantages arising from the
partitioned model, in addition to the improved performance
obtained for the overlay and location-based operations:

1. The partitioned model is particularly suitable for
   parallel I/O and processing [4, 9, 25].

2. The partitioned approach enables the effective trans-
   mission of data; independent pieces are easier to
   transmit than large interdependent ones. This capabil-
   ity is critical in a network environment.

In summary, although the arguments given are far from
conclusive (indeed, they depend on a particular but represen-
tative choice of the vector, seamless and partitioned
models and on the sizes and organization of the tiles), we
believe that there are strong reasons for continued research
into large partitioned spatial databases. This is especially
ture in view of the recent advances in parallel computing
technology for which the partitioned approach appears to be
particularly suited.

References

1. M. S. Bundock, "An Integrated DBMS Approach to
   Geographical Information Systems",  Proc. Auto Carto
   8, Baltimore, Maryland, 1987, pp. 292-301

2. P. A. Burrough,  Principles of Geographical Information
   Systems for Land Resources Assessment,

3. A. U. Frank, "Requirements for Database Systems
   Suitable to Manage Large Spatial Databases", Proc.
   International Symposium on Spatial Data Handling,
   Vol 1, Zurich, Switzerland, pp. 38-60, 1984

4. W. R. Franklin, C. Narayanaswami, M. Kankanhalli, D.
   Sun, M. Zhou, P. Y. Wu, "Uniform Grids: A Technique
   for Intersection Detection on Serial and Parallel
   Machines",  Proc. Auto Carto 9, Baltimore, Maryland,
   pp. 100-109, 1989

5. I. Gargantini, "An Effective Way to Represent
   Quadtrees",  Communications of the ACM, Vol. 25,
   No. 12, 1982, pp. 905-910

6. M. Goodchild, "Optimal Tiling for Large Cartographic
   Databases",  Proc. Auto Carto 9, Baltimore, Maryland,
   pp. 444-451, 1989

7. S. C. Guptill, "Speculation on Seamless, Scaleless
   Cartographic Data Bases",  Proc. Auto Carto 9, Balti-
   more, Maryland, pp. 436-443, 1989

8. A. Gutman, "R-Trees: A Dynamic Index Structure For
   Spatial Searching", Proc. ACM SIGMOD Conference
   on Management of Data, Boston, Mass., pp. 47-57,
   1984

9. R. G. Healey, G. B. Desa, "Transputer Based Parallel
   Processing for GIS Analysis: Problems and Potentiali-
   ties", Proc. Auto Carto 9, Baltimore, Maryland, pp. 90-
   99, 1989

    Cartographic Data Collections - An Empirical Study",
    Proc. Auto Carto 9, Baltimore, Maryland, pp. 416-425,
    1989

11. R. A. Lorie, A. Meier, Using A Relational DBMS for
    Geographical Databases, Computer Science Re-
    search Report 3848 (43915), IBM Research Labora-
    tory, San Jose, CA, 1983

12. R. W. Marx, "The TIGER System: Automating the
    Geographic Structure of the United States Census",
    Government Publications Review, 13, pp. 181-201,
    1986

13. D. McGregor, "Geographic Information System
    Trends",  Proc. Third International GIS/LIS Confer-
    ence, San Antonio, Texas, pp. 915-921, 1988


22. L. Vanzella, *Classification of Data Structures for Thematic Data*, Dept. of Computing Science TR 88-14, University of Alberta, Edmonton, Canada, 1988


Data Issues and Remote Sensing

Session F6
Data Issues

Using CAD Data in a GIS

Automated GIS Data Conversion Technology with Case Studies

GIS for Forest Information Management: A Case Study
Using CAD Data in a GIS

Carl N. Reed, III
President
Genasys, Inc.
2629 Redwing Rd., Suite 330
Ft. Collins, Colorado
0526
USA

Abstract
The most costly component of implementing a GIS capability within an organization is data conversion. Consequently, the ability to use existing digital data that may already exist in CAD systems or in other GIS’s is extremely important. Unfortunately, there can be numerous (and very frustrating) problems encountered that make digital data conversion a nightmare. To avoid these conversion problems, the GIS users can follow a series of logical steps and rules for evaluating and converting existing digital map data. These steps include defining a data definitions matrix, defining data organization rules, and defining symbolization rules before any data are either digitized or converted. They also included rules of conversion that relate to coordinate systems, transforming non-topology based vector data into topological data structures, and attaching attributes to the geographic features in the GIS database. If a set of rules and procedures are followed, then the conversion of existing digital map data can realize major cost savings and get the new GIS user into production in a much shorter time frame.
The Problem
Upon procuring a GIS, many organizations are faced with the dilemma of how to populate the database. As an example, consider a City as the new GIS user. Typically, they have map source documents scattered throughout the organization. Each department may have its own map set. These documents are at different scales, in different condition, and so on. Further, many of these documents will have no geographic control on them. To make matters more interesting, one group has digitized their particular map set with a PC-based CAD system and would like to put these data into the GIS. At the same time, the users are struggling with the decision as to whether digitizing should occur in-house or under contract with some conversion house. Then, just as the GIS users group begins to develop a solid plan for developing their digital map database, someone discovers that a sister organization has digitized the base map for the City. However, these data currently reside in a CAD environment.

Obviously, solving the problem of data entry can appear to be a somewhat daunting effort. As much or more effort needs to go into the definition and prototype development of the digital map database as does the actual procurement of the GIS hardware and software. As can be seen from the above example, there are a variety of issues that add complexity to the process. One of these is how to integrate existing and future digital map data that was compiled using CAD system. In the example, there are three possible sources of CAD data. The first is the existing in-house PC CAD data. The second possible source will occur if the City decides to have contract digitizing and the award is to a conversion house that has only CAD systems. The third sources is the sister organization with the Intergraph system.

Typically, there is considerable pressure to use CAD data because it will “get us operational more quickly so that we can show management that the system can produce useful products”. The watch phrase is, “Look before you leap”!

Understanding CAD Data
Before beginning a discussion of how CAD data should be processed for integration into a GIS database, it is important to understand several key concepts regarding how data should be stored within a GIS. These concepts are not necessary components for building a CAD database. However, if they are not used in building the CAD database, the CAD data becomes much more difficult, if not impossible, to use in a GIS database.

Projections/Coordinate Systems
A GIS requires that data be stored using some projected coordinated system. A very common system for municipalities in the United States is the State Plane Coordinate System. There are many reasons for storing the coordinate data in a consistent and definable manner. These reasons include area computations, length computations, edge match between map sheets, the need to build seamless/continuous layers, and map plotting to scale with a reference grid. There may also be legal considerations.

In a CAD environment, data can be collected in any arbitrary rectilinear coordinate system that preserves relationships within a drawing. Thus, a map could be digitized in inches and stored in the CAD drawing database in inches. Whenever that map is plotted in the CAD system, it will look correct. However, if those same data are exported into a GIS, they will, when plotted, bear absolutely no relationship to map data stored in the GIS. The best case scenario for data not in a projected system is that it will have to be “rubber sheeted” to fit the GIS database. There are several problems with this. First, it is one more processing step. Second, it will introduce error into the data that was not originally there. Third, it will require an additional QC step. Further, if the CAD system has COGO tools and these were used to enter survey data, then the GIS must have tools that allows for “fixing” monumented points. This is important because rubber sheeting is a global transformation which will shift all coordinates.

Edge Match
A good GIS is designed to allow the users to develop a seamless map environment. This means that the software and the database structure support the ability to edgematch data within a layer across map source document boundaries. If a CAD system has been used to collect map data on a sheet by sheet basis, in all probability there has been no edge match performed. Therefore, to use these data, an edge match process will have to be performed and the results validated. Sometimes, in the CAD environment, many sheets will have been digitized into one drawing file. In this case, edge match has been performed. However, there is now the alternative problem. Most GIS partition the map data for better processing efficiency. Therefore, data in which all maps are in one file will have to be partitioned into smaller units. There is still an additional processing step.

Topology
In a GIS, a data storage philosophy called topology is used. In this environment, each intersection or node contains information about what lines are associated with it. Further,