The GIS Primer

An Introduction to Geographic Information Systems

David J. Buckley
Principal GIS Consultant
Innovative GIS Solutions, Inc.

February 1997

GIS Solutions Inc.
INNOVATIVE
GIS with a vision...

Table of Contents

APPENDIX A - Glossary of Terms 106

Although the GIS Primer was developed more than 10 years ago I continue to receive requests for the document. Originally developed in the late 1980's (thanks to the support of the Alberta government), the GIS Primer was intended to be the workbook for an Introduction to GIS workshop that I developed. At that time there were only 2 GIS textbooks available and neither of them provided a simple overview of GIS targeted towards natural resource professionals.

In our time at Innovative GIS in the 1990's we continued to update the Primer and provide it to those who asked for it. My colleague, Dr. Joseph Berry, has always hosted the HTML version on his web site however I recently found it on a backup and thought since it just won't die, it is about time to provide it in a more useable format. So, here is the latest version I have in PDF format (circa 1997)

Many of the technical references in the Primer are dated, however the core content remains valid. I have provided the document "as is" with original content, examples and references. With the plethora of GIS texts now available I hope the GIS Primer still continues to be helpful to those who are new to GIS. The author can be reached at *dbuckley14@live.com*

David Buckley, January 2008

Preface

This book is a result of several years experience in applying **G**eographic **I**nformation **S**ystems (GIS) technology to typical forestry, natural resource management, and environmental problems. Many of the terms and concepts introduced in this book have been compiled from a variety of different sources. These include some of the prevalent text books in the field, but most have been firmly implanted from practical experience. GIS is in essence, an applied science. The more *hands-on* you have the more entrenched the methodologies and techniques become for effectively applying the technology.

The sample applications presented in the GIS Primer have been compiled from my experiences representing real operational requirements of selected IGIS clients, past and present. These sources include provincial and federal natural resource agencies in Canada and the United States, and private forestry companies across North America. The models and/or formulas used in the sample applications have been simplified in an effort to focus on procedures and techniques for applying GIS. The sample applications are presented to provide simple operational examples of fundamental GIS analysis techniques.

Annual updates are undertaken to this book to reflect the dynamic nature of the commercial GIS marketplace, and to reflect my understanding of *what GIS really is*. The ever increasing rate of technological change prompts me to update the Primer to reflect the new and exciting capabilities that GIS technology provides. For the novice reader, don't be overwhelmed with this change. GIS has changed so drastically since this book was originally written in 1988; but in so many ways remains the same. Changing technology has provided many more capabilities and possibilities, often beyond the comprehension of the average user. Nonetheless, ultimately the tool is only as good as the craftsman using it. Consistently it is the imagination and creativity of the *domain specialist* using GIS that makes GIS work, not the new technological capabilities or increased performance. I believe that while the GIS vendor community, hardware and software vendors alike, provide us with newer, better and faster technological tools, it is in the end, the domain specialists applying the tool that define *state-of-the-art*. The heartbeat of GIS still lies in the field and district offices, the logging divisions, the engineering offices, and the with the small GIS entrepreneurs in offices everywhere.

Many thanks to my colleagues, past and present, who have supported me and worked with me over the years. This book reflects our collective experiences and knowledge. A special thanks to my partners at IGIS who share a common vision of the role of GIS in today's society. And lastly, to Joe Berry, my GIS partner in crime, who continues to amaze me with his insight and genuine love for all things GISsy. Many of the figures and diagrams in the Primer are graciously provided by Joe from his *gCON Software* set.1 Many thanks for all your support.

The GIS Primer is provided as the workbook and reference document for all our introductory GIS education workshops. Permission is hereby granted to copy and distribute this document for educational purposes. I request that you provide appropriate references when portions of this document are utilized. Thanks.

David J. Buckley
Fort Collins, Colorado
February 1997

1 The gCON Software is a set of 19 digital slide shows that present illustrations of basic GIS concepts, analytical techniques, modeling procedures and applications presented in Dr. Berry's books "Beyond Mapping: Concepts, Algorthims, and Issues in GIS" and "Spatial Reasoning for Effective GIS". They can be purchased from GIS World Inc. , Tel: 970-223-4848, FAX:970-223-5700 or E-mail: jberry@innovativegis.com.

Purpose of the Book

The purpose of this book is to provide an overview of the issues and requirements for implementing and applying geographic information systems technology.

The book is intended to provide resource specialists and managers with an exposure to GIS technology and terminology, so that papers, applications, and exhibits on GIS will be more understandable and useful in the future. This book presents a substantial amount of GIS theory, and a whole bunch of buzzwords, but is also supplemented with relevant forestry based applications to illustrate the concepts and techniques for applying GIS technology.

The focus of this book is on practical issues concerned with the implementation and application of GIS technology. This book is directed at the resource specialist and/or manager who has had limited exposure to GIS. It is intended purely as an introductory text with an emphasis on identifying and clearly illustrating primitive concepts in geographic information system applications.

Hopefully, with this approach in mind the confusion and mystique that often surrounds GIS can be clarified, resulting in an understandable framework for the successful evaluation, implementation, and application of GIS technology.

This book is based purely on practical experience of the author and in no way should be construed as a comprehensive statement on GIS. References are provided at the back of the book for further reading and detail.

Important points or quotes are identified throughout the book by **bold** text. In some cases a **box** is used to highlight a key quote or phase. *Italics* are used throughout the text to identify technical terms that imply something other that their conventional meaning. Many of these are defined in the Glossary found in the Appendix A.

Any comments or suggestions on the content or format of this book are welcome. Please send any comments to “buck@innovativegis.com”.

Enjoy !

1.0 The Nature of Geographic Information

This chapter reviews the basic fundamentals of geographic data and information. The focus is on understanding the basic structure of geographic data, and how issues of accuracy, error, and quality are paramount to properly using GIS technology. The establishment of a robust database is the cornerstone of a successful GIS.

1.1 Maps and Spatial Information

The main method of identifying and representing the location of geographic features on the landscape is a **map**. A map is a graphic representation of where features are, explicitly and relative to one another. A map is composed of different geographic features represented as either points, lines, and/or areas. Each feature is defined both by its location in space (with reference to a coordinate system), and by its characteristics (typically referred to as attributes). Quite simply, a map is a model of the real world.

The **map legend** is the key linking the attributes to the geographic features. Attributes, e.g. such as the species for a forest stand, are typically represented graphically by use of different symbology and/or color. For GIS, attributes need to be coded in a form in which they can be used for data analysis (Burrough, 1986). This implies loading the attribute data into a database system and *linking* it to the graphic features.

For geographic data, often referred to as *spatial data*, features are usually referenced in a coordinate system that models a location on the earth's surface. The coordinate system may be of a variety of types. For natural resource applications the most common are :

- **geographic coordinates** such as latitude and longitude, e.g. 56°27'40" and 116°11'25". These are usually referred by degrees, minutes, and seconds. Geographic coordinates can also be identified as decimal degrees, e.g. 54.65°.

- **a map projection**, e.g. Universe Transverse Mercator (UTM) where coordinates are measured in metres, e.g. 545,000.000 and 6,453,254.000 normally reference to a central meridian. Eastings refer to X coordinates while Northings refer to Y coordinates.

- **a legal survey description**, e.g. Meridian, Township, Range such as the Alberta Township System, e.g. Township 075 Range 10 West of 4th Meridian.

Geographic data is distinguished from *attribute data* in that it is referenced spatially by a coordinate system, e.g. it has a spatial extent. Natural resource applications commonly use a Legal Survey system, e.g. the Alberta Township System (ATS), which identifies feature's locations as being a Meridian, Township, and a Range; or a projection such as the UTM coordinate system which identifies features by an Easting coordinate (X) and a Northing coordinate (Y) in a particular UTM zone.

In the UTM projection the area of the earth between 80 degrees North and 80 degrees South latitude is divided into north-south columns 6 degrees of longitude wide called *zones*. These are numbered 1 to 60 eastward beginning at the 180th meridian. Within each zone the central meridian is given an Easting value of 500,000 metres. The equator is designated as having a Northing value of 0 for northern hemisphere coordinates. Coordinates are recorded relative to the central meridian in metres in a particular zone. The basis of the UTM projection defines that the coordinates are duplicated within each UTM zone. Accordingly, use of the UTM projection is only appropriate for certain spatial extents and scales of data. It is not appropriate to use this projection if your area of interest crosses UTM zone boundaries.

For example, the province of Alberta is located in UTM zones 11 and 12. The central meridian for zone 11 is 117° longitude. The central meridian for zone 12 is 111° longitude. The UTM coordinate system is the most widely used projection in the mapping industry and consequently is becoming an de facto standard for use with geographic information systems. This is particularly true for regional data in Canada. The State Plane coordinate system is widely used in the United States.

Maps are the traditional method of storing and displaying geographic information. **A map portrays 3 kinds of information about geographic features. The :**

- **Location and extent of the feature;**
- **Attributes (characteristics) of the feature**; and
- **Relationship of the feature to other features.**

Geography has often been described as the study of **why what is where**. This description is quite appropriate when considering the three kinds of information that are portrayed by the traditional map;

- the **location and extent** of a feature is identified explicitly by reference to a coordinate system representing the earth's surface. This is **where** a feature is.
- the **attributes** of a feature describe or characterize the feature. This is **what** the feature is.
- The relationship of a feature to other features is implied from the location and attributes of all features. Relationships can be defined explicitly, e.g. roads connecting towns, regions adjacent to one another, or implicitly, e.g. close to, far from, similar to, etc. Implicit relationships are interpreted according to the knowledge that we have about the natural world. Relationships are described as **how** or **why** a feature is.

.

The geographic information system distinguishes between the *spatial* and *attribute* aspect of geographic features.

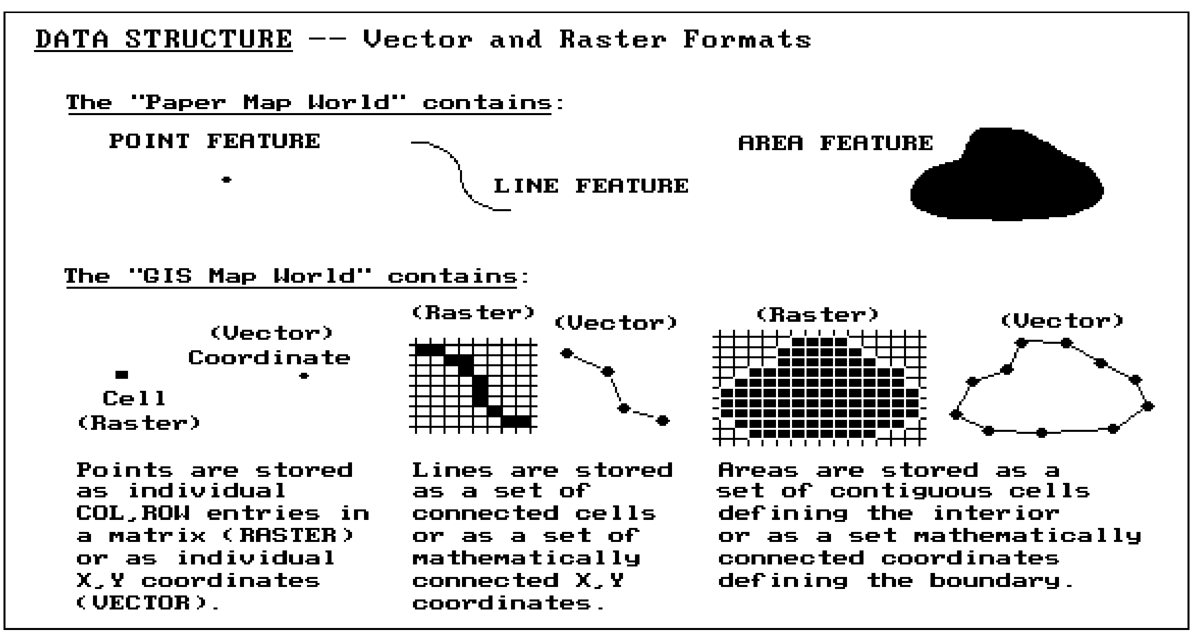
The identification of relationships between features, within a common theme or across different themes, is the primary function of a GIS.

1.2 Characterizing Geographic Features

All geographic features on the earth's surface can be characterized and defined as one of three basic feature types. These are **points, lines,** and **areas**.

- **Point** data exists when a feature is associated with a single location in space. Examples of point features include a fire lookout tower, an oil well or gas activity site, and a weather station.
- **Linear** data exists when a feature's location is described by a string of spatial coordinates. Examples of linear data include rivers, roads, pipelines, etc.
- **Areal** data exists when a feature is described by a closed string of spatial coordinates. An area feature is commonly referred to as a **polygon**. Polygonal data is the most common type of data in natural resource applications. Examples of polygonal data include forest stands, soil classification areas, administrative boundaries, and climate zones. Most polygon data is considered to be homogeneous in nature and thus is consistent throughout.

Every geographic phenomenon can in principle be represented by either a point, a line, and/or an area.



Commonly, an *identifier* accompanies all types of geographic features. This description or identifier is referred to as a **label**. Labels distinguish geographic features of the same type, e.g. forest stands, from one another. Labels can be in the form of a name, e.g. "Lake Louise", a description, e.g. "WELL" or a unique number, e.g. "123". Forest stand numbers are examples of polygon labels. Each label is unique and provides the mechanism for linking the feature to a set of descriptive characteristics, referred to as attribute data.

It is important to note that geographic features and the symbology used to represent them, e.g. point, line, or polygon, are dependant on the *graphic scale* (*map scale*) of the data. Some features can be represented by point symbology at a small scale, e.g. villages on a 1:1,000,000 map, and by areal symbology at a larger scale, e.g. villages on a 1:10 ,000 map. Accordingly, the accuracy of the feature's location is often *fuzzier* at a smaller scale than a larger scale. The generalization of features is an inherent characteristic of data presented at a smaller scale.

Data can always be generalized to a smaller scale,
but detail CANNOT be created !

Remember, as the scale of a map increases, e.g. 1:15,000 to 1:100,000, the relative size of the features **decrease** and the following may occur :

- Some features may disappear, e.g. features such as ponds, hamlets, and lakes, become indistinguishable as a feature and are eliminated.;
- Features change from areas to lines or to points, e.g. a village or town represented by a polygon at 1:15,000 may change to point symbology at a 1:100,000 scale.;
- Features change in shape, e.g. boundaries become less detailed and more generalized.; and
- Some features may appear, e.g. features such as climate zones may be indistinguishable at a large scale (1:15,000) but the full extent of the zone becomes evident at a smaller scale (1:1,000,000).

Accordingly, the use of data from vastly different scales will result in many inconsistencies between the number of features and their type.

The use and comparison of geographic data from vastly different source scales is totally inappropriate and can lead to significant error in geographic data processing.

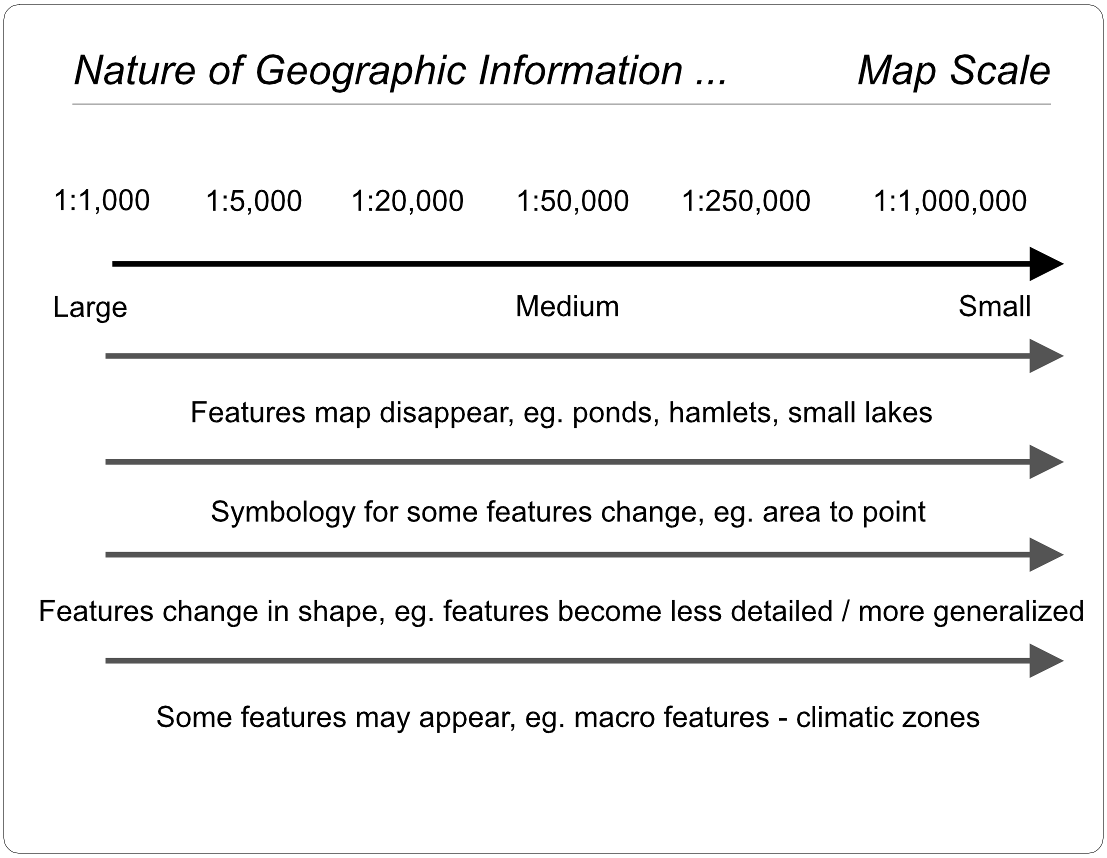


Figure 1. This represents a summary of map scale issues.

1.3 Data Accuracy and Quality

The quality of data sources for GIS processing is becoming an ever increasing concern among application personnel. With the influx of GIS software on the commercial market and the accelerating application of GIS technology to problem solving and decision making roles, the quality and reliability of GIS products is coming under closer scrutiny. Much concern has been raised as to the relative error that may be inherent in GIS processing methodologies. While research is ongoing, and no finite standards have yet been adopted in the commercial GIS marketplace, several practical recommendations have been identified which help to locate possible error sources, and define the quality of data. The following review of data quality focuses on three distinct components, data accuracy, quality, and error.

1.3.1 Accuracy

The fundamental issue with respect to data is *accuracy*. **Accuracy is the closeness of results of observations to the true values or values accepted as being true**. This implies that observations of most spatial phenomena are usually only considered to estimates of the true value. The difference between observed and true (or accepted as being true) values indicates the accuracy of the observations.[2]

[2] An excellent review of accuracy and quality is provided by Hunter and Williamson, 1990.

Basically two types of accuracy exist. These are **positional** and **attribute** accuracy.[3] **Positional accuracy** is the expected deviance in the geographic location of an object from its true ground position. This is what we commonly think of when the term accuracy is discussed. There are two components to positional accuracy. These are *relative* and *absolute* accuracy. *Absolute accuracy* concerns the accuracy of data elements with respect to a coordinate scheme, e.g. UTM. *Relative accuracy* concerns the positioning of map features relative to one another.

Often relative accuracy is of greater concern than absolute accuracy. For example, most GIS users can live with the fact that their survey township coordinates do not coincide exactly with the survey fabric, however, the absence of one or two parcels from a tax map can have immediate and costly consequences.[4]

Attribute accuracy is equally as important as positional accuracy. It also reflects estimates of the truth. Interpreting and depicting boundaries and characteristics for forest stands or soil polygons can be exceedingly difficult and subjective. Most resource specialists will attest to this fact. Accordingly, the degree of homogeneity found within such mapped boundaries is not nearly as high in reality as it would appear to be on most maps.[5]

1.3.2 Quality

Quality can simply be defined as the fitness for use for a specific data set. Data that is appropriate for use with one application may not be fit for use with another. It is fully dependant on the scale, accuracy, and extent of the data set, as well as the quality of other data sets to be used. The recent U.S. Spatial Data Transfer Standard (SDTS) identifies five components to data quality definitions. These are :

- Lineage
- Positional Accuracy
- Attribute Accuracy
- Logical Consistency
- Completeness

1. Lineage

The lineage of data is concerned with historical and compilation aspects of the data such as the :

- source of the data;
- content of the data;
- data capture specifications;
- geographic coverage of the data;

[3] Aronoff, 1989.

[4] Hunter, Williamson, 1990.

[5] Hunter, Williamson, 1990.

- compilation method of the data, e.g. digitizing versus scanned;
- transformation methods applied to the data; and
- the use of an pertinent algorithms during compilation, e.g. linear simplification, feature generalization.

2. Positional Accuracy

The identification of positional accuracy is important. This includes consideration of inherent error (source error) and operational error (introduced error). A more detailed review is provided in the next section.

3. Attribute Accuracy

Consideration of the accuracy of attributes also helps to define the quality of the data. This quality component concerns the identification of the reliability, or level of purity (homogeneity), in a data set.

4. Logical Consistency

This component is concerned with determining the faithfulness of the data structure for a data set. This typically involves spatial data inconsistencies such as incorrect line intersections, duplicate lines or boundaries, or gaps in lines. These are referred to as spatial or *topological* errors.

5. Completeness

The final quality component involves a statement about the completeness of the data set. This includes consideration of *holes* in the data, unclassified areas, and any compilation procedures that may have caused data to be eliminated.

The ease with which geographic data in a GIS can be used at any scale highlights the importance of detailed data quality information. Although a data set may not have a specific scale once it is loaded into the GIS database, it was produced with levels of accuracy and resolution that make it appropriate for use only at certain scales, and in combination with data of similar scales.

1.3.3 Error

Two sources of error, **inherent** and **operational**, contribute to the reduction in quality of the products that are generated by geographic information systems. **Inherent error** is the error present in source documents and data. **Operational error** is the amount of error produced through the data capture and manipulation functions of a GIS.[6] Possible sources of operational errors include :

- mislabelling of areas on thematic maps;
- misplacement of horizontal (positional) boundaries;
- human error in digitizing;
- classification error;
- GIS algorithm inaccuracies; and
- human bias.

While error will always exist in any scientific process, the aim within GIS processing should be to identify existing error in data sources and minimize the amount of error added during processing. Because of cost constraints it is often more appropriate to manage error than attempt to eliminate it. There is a trade-off between reducing the level of error in a data base and the cost to create and maintain the database.

An awareness of the error status of different data sets will allow a user to make a subjective statement on the quality and reliability of a product derived from GIS processing.

The validity of any decisions based on a GIS product is directly related to the quality and *reliability rating* of the product.

Depending upon the level of error inherent in the source data, and the error operationally produced through data capture and manipulation, GIS products may possess significant amounts of error.[7]

One of the major problems currently existing within GIS is the **aura of accuracy** surrounding digital geographic data. Often hardcopy map sources include a map *reliability rating* or *confidence rating* in the map legend. This rating helps the user in determining the *fitness for use* for the map. However, rarely is this information encoded in the digital conversion process.

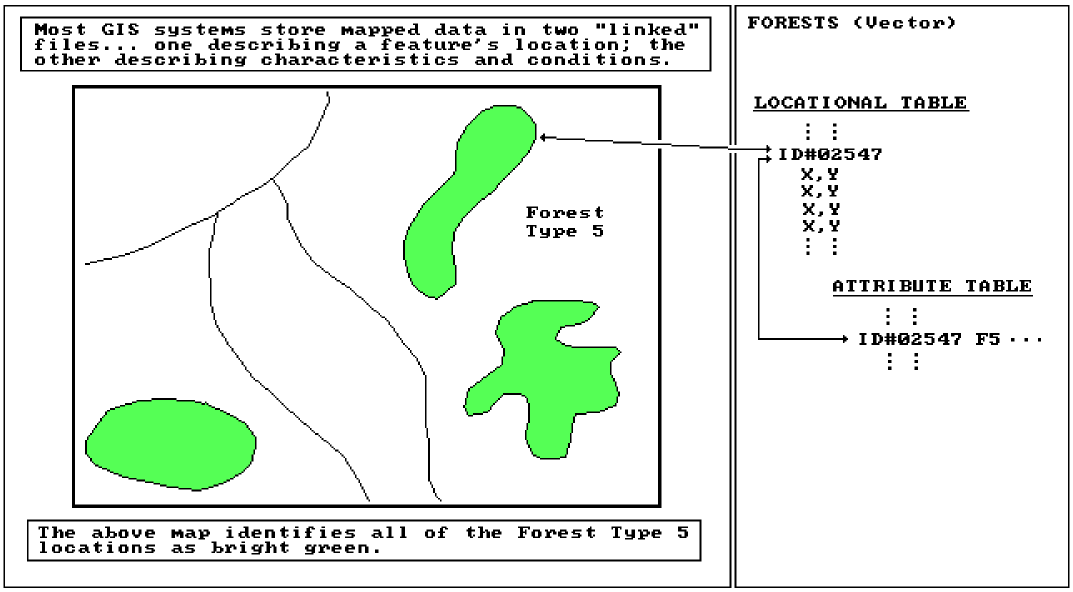
Often because GIS data is in digital form and can be represented with a high *precision* it is

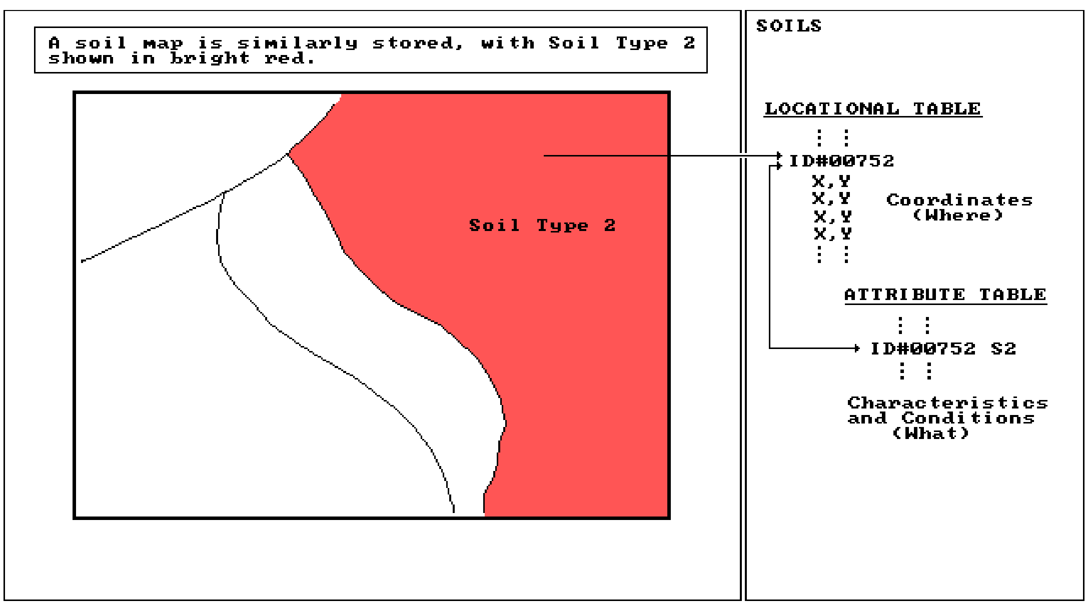
[6] Walsh et all, 1987.

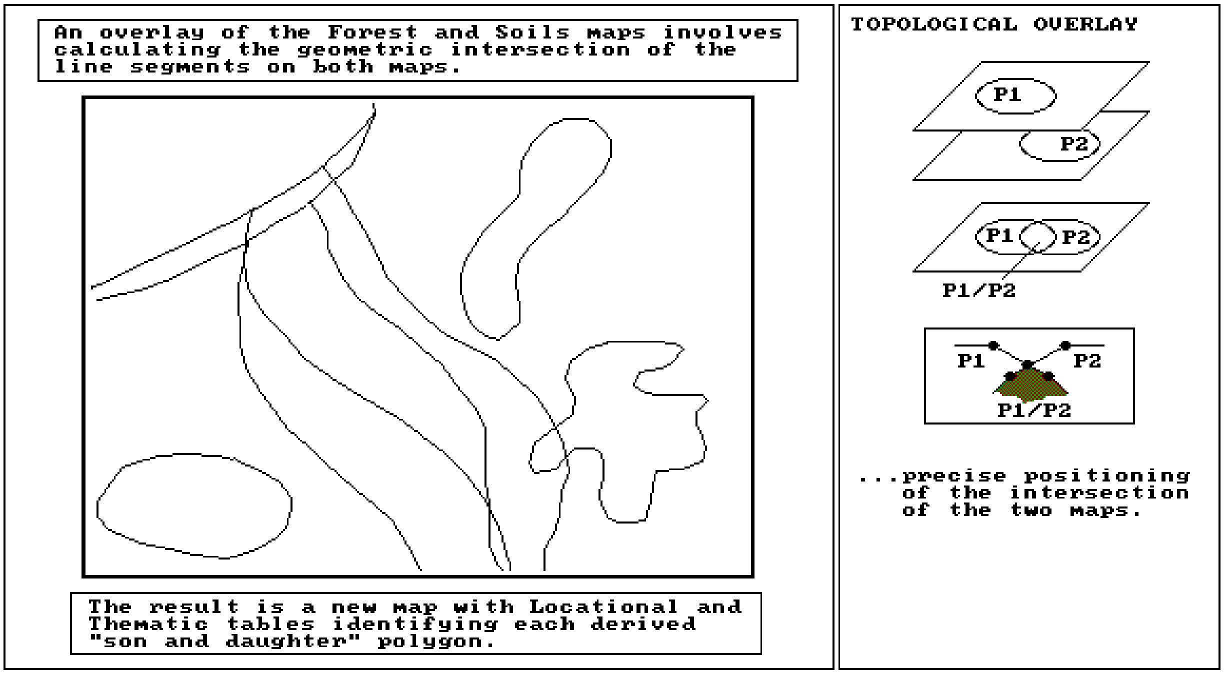
[7] Walsh et al, 1987.

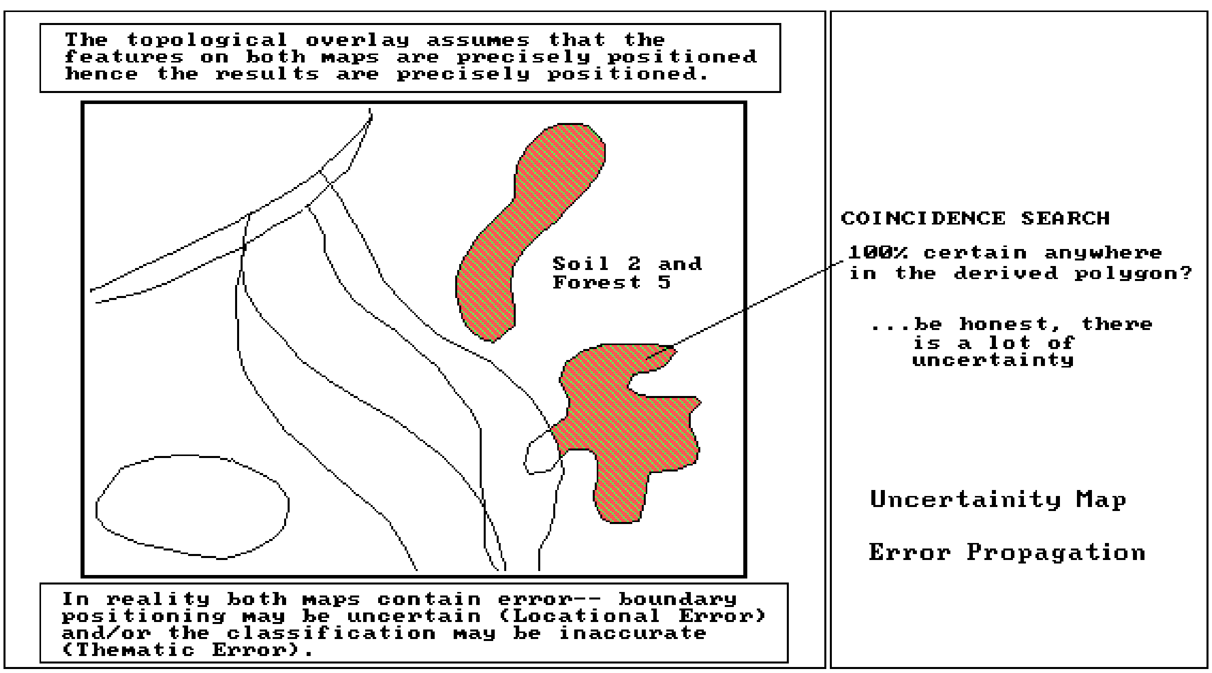
considered to be totally *accurate*. In reality, a *buffer* exists around each feature which represents the actual positional location of the feature. For example, data captured at the 1:20,000 scale commonly has a positional accuracy of +/- 20 metres. This means the actual location of features may vary 20 metres in either direction from the identified position of the feature on the map. Considering that the use of GIS commonly involves the integration of several data sets, usually at different scales and quality, one can easily see how errors can be propagated during processing.

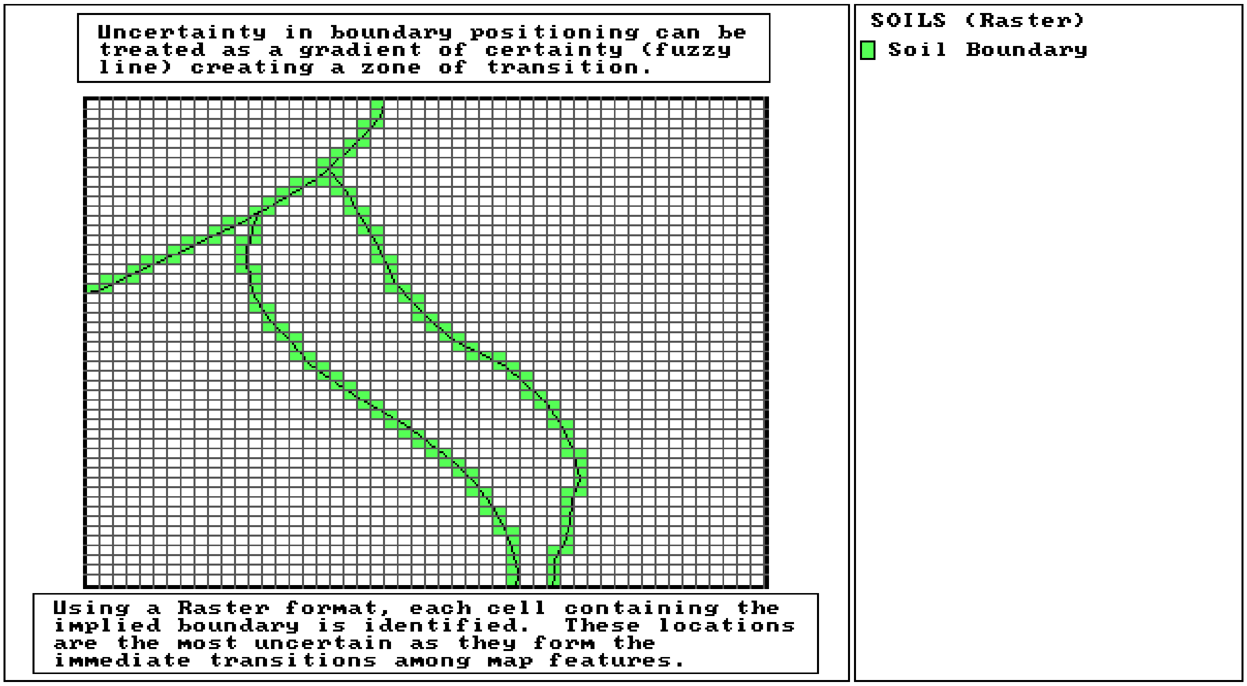
The following series of slides illustrate how error can be propagated when spatial data layers are combined by presenting the *zone of uncertainty* that actually surround data of a specific map scale. The example illustrates how, using raster modeling techniques, *maps of uncertainty* can be generated that reflect a level of confidence in data combined using GIS.

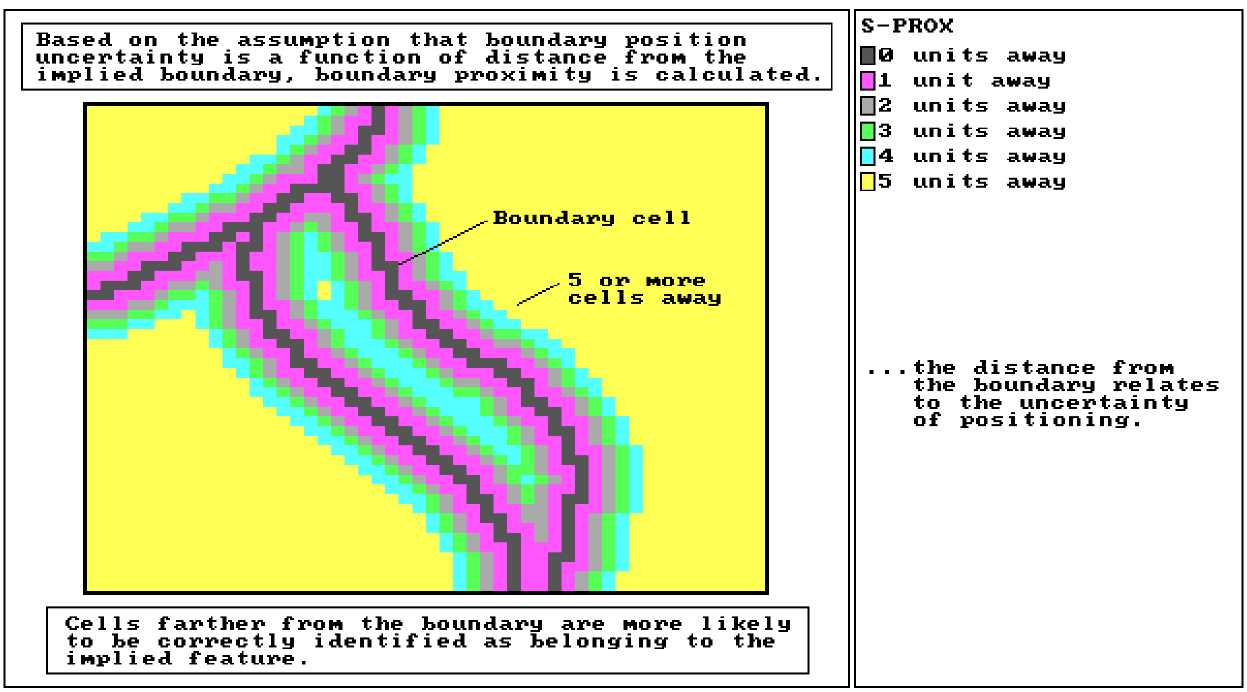


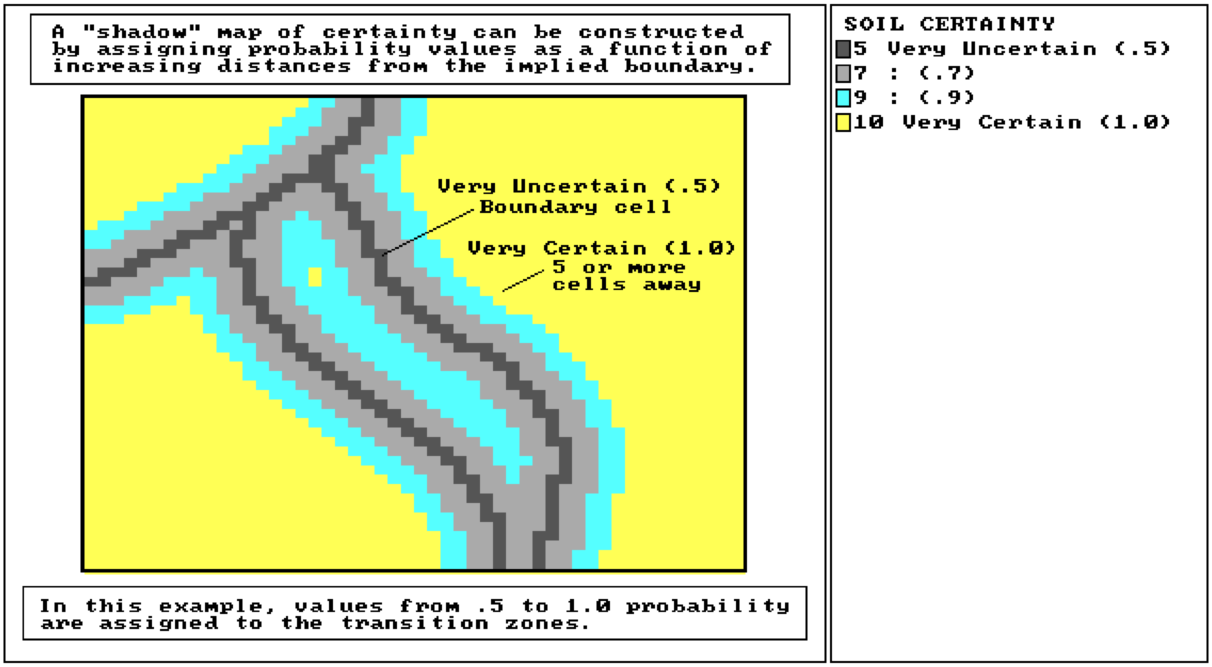


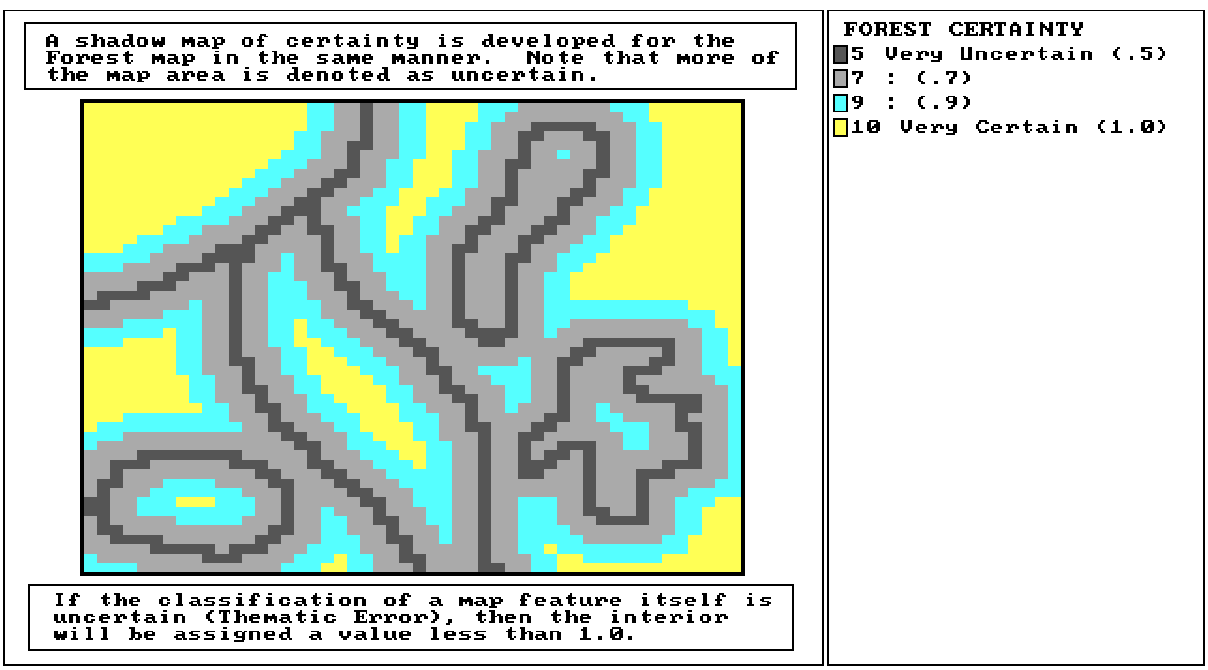

An overlay of the Forest and Soils maps involves calculating the geometric intersection of the line segments on both maps.
The result is a new map with Locational and Thematic tables identifying each derived "son and daughter" polygon.
TOPOLOGICAL OVERLAY
P1
P2
P1
P2
P1/P2
P1
P2
P1/P2
...precise positioning of the intersection of the two maps.

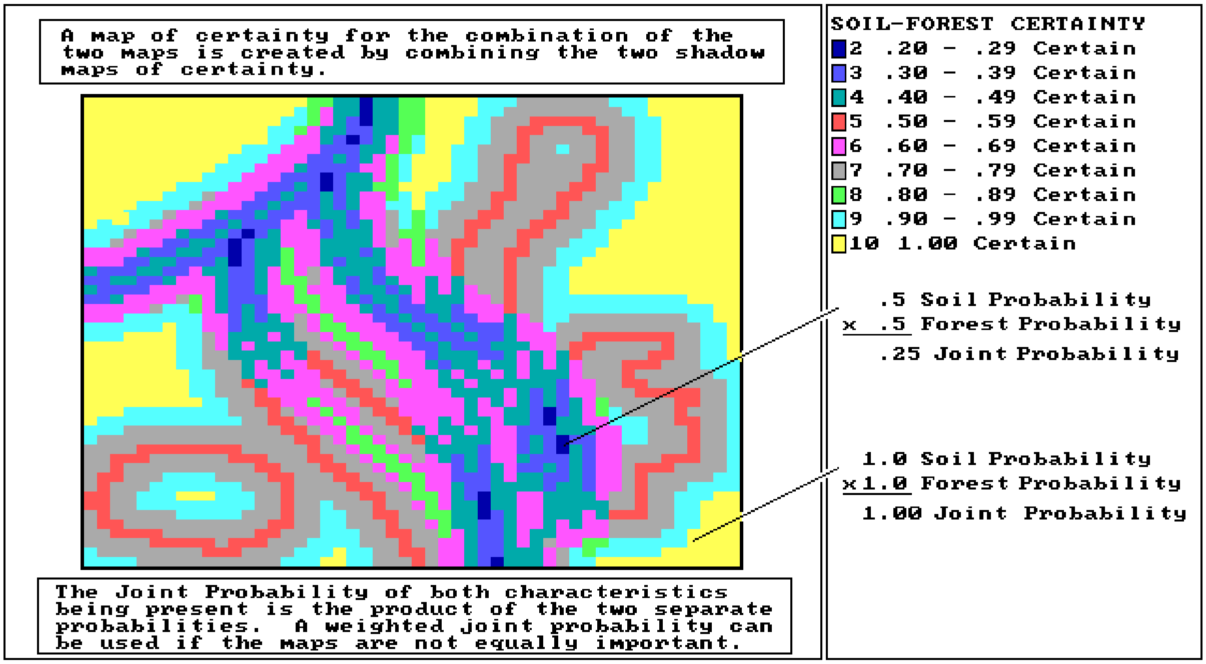

The topological overlay assumes that the features on both maps are precisely positioned hence the results are precisely positioned.
Soil 2 and Forest 5
In reality both maps contain error-- boundary positioning may be uncertain (Locational Error) and/or the classification may be inaccurate (Thematic Error).
COINCIDENCE SEARCH
100% certain anywhere in the derived polygon?
...be honest, there is a lot of uncertainty
Uncertainity Map
Error Propagation

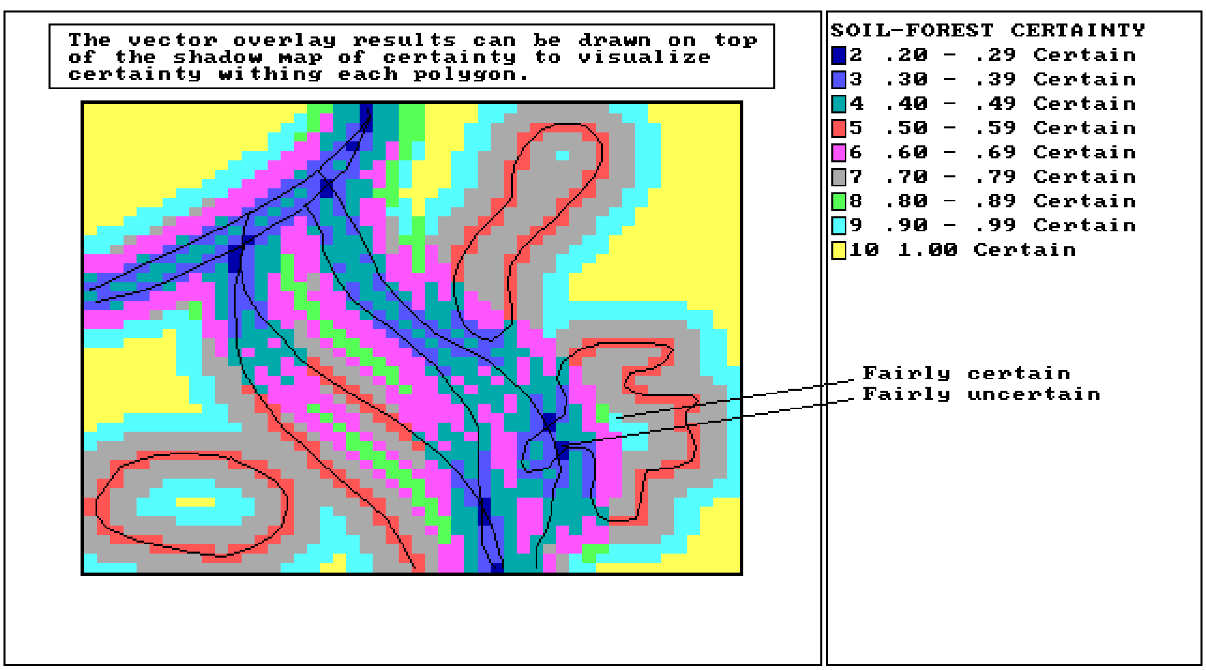

Uncertainty in boundary positioning can be treated as a gradient of certainty (fuzzy line) creating a zone of transition.
SOILS (Raster)
Soil Boundary
Using a Raster format, each cell containing the implied boundary is identified. These locations are the most uncertain as they form the immediate transitions among map features.

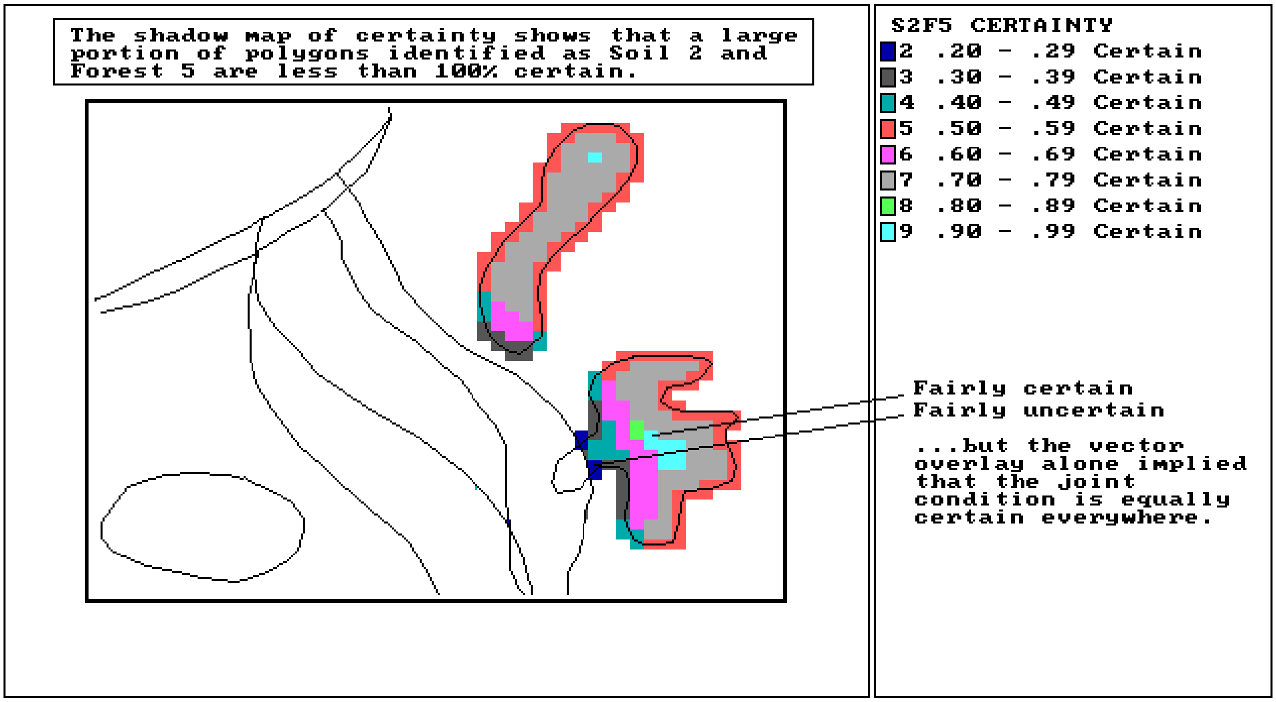

Based on the assumption that boundary position uncertainty is a function of distance from the implied boundary, boundary proximity is calculated.
S-PROX
0 units away
1 unit away
2 units away
3 units away
4 units away
5 units away
Boundary cell
5 or more cells away
...the distance from the boundary relates to the uncertainty of positioning.
Cells farther from the boundary are more likely to be correctly identified as belonging to the implied feature.

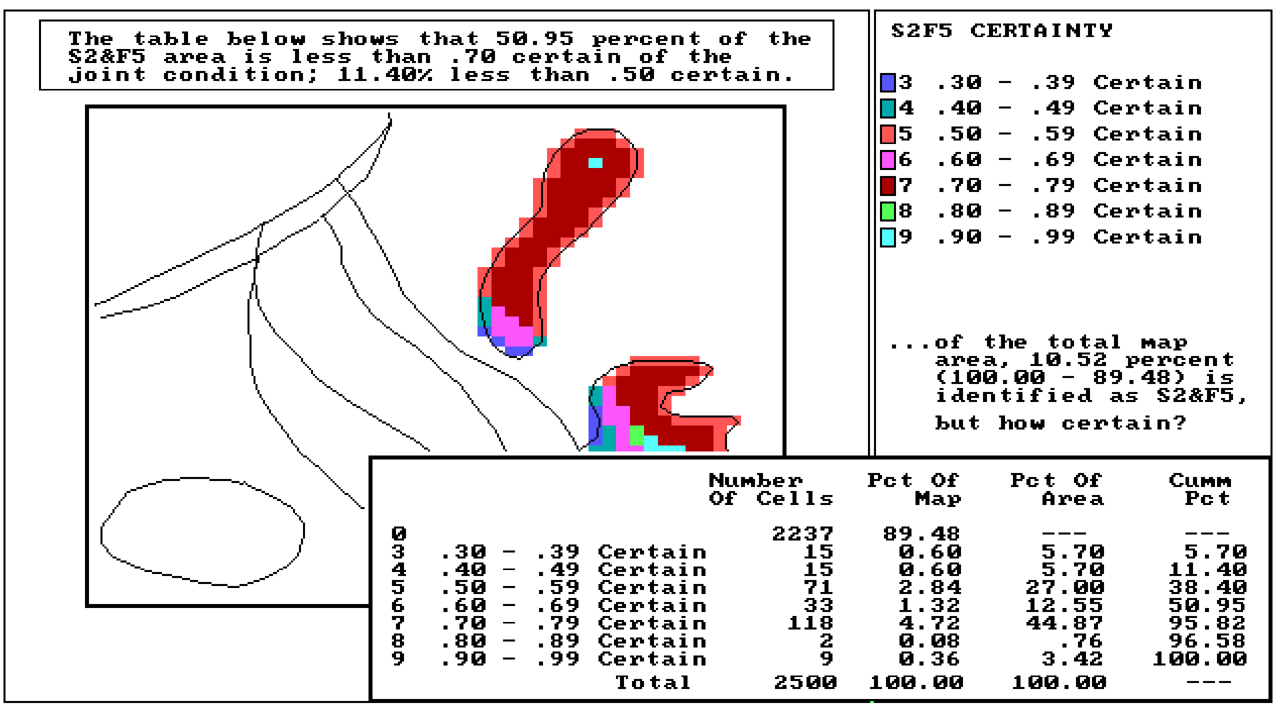

A "shadow" map of certainty can be constructed by assigning probability values as a function of increasing distances from the implied boundary.
Very Uncertain (.5)
Boundary cell
Very Certain (1.0)
5 or more cells away
In this example, values from .5 to 1.0 probability are assigned to the transition zones.
SOIL CERTAINTY
5 Very Uncertain (.5)
7 : (.7)
9 : (.9)
10 Very Certain (1.0)


A shadow map of certainty is developed for the Forest map in the same manner. Note that more of the map area is denoted as uncertain.
If the classification of a map feature itself is uncertain (Thematic Error), then the interior will be assigned a value less than 1.0.
FOREST CERTAINTY
5 Very Uncertain (.5)
7 : (.7)
9 : (.9)
10 Very Certain (1.0)


A map of certainty for the combination of the two maps is created by combining the two shadow maps of certainty.
SOIL-FOREST CERTAINTY
2 .20 - .29 Certain
3 .30 - .39 Certain
4 .40 - .49 Certain
5 .50 - .59 Certain
6 .60 - .69 Certain
7 .70 - .79 Certain
8 .80 - .89 Certain
9 .90 - .99 Certain
10 1.00 Certain
.5 Soil Probability
x .5 Forest Probability
.25 Joint Probability
1.0 Soil Probability
x 1.0 Forest Probability
1.00 Joint Probability
The Joint Probability of both characteristics being present is the product of the two separate probabilities. A weighted joint probability can be used if the maps are not equally important.


The vector overlay results can be drawn on top of the shadow map of certainty to visualize certainty withing each polygon.
SOIL-FOREST CERTAINTY
2 .20 - .29 Certain
3 .30 - .39 Certain
4 .40 - .49 Certain
5 .50 - .59 Certain
6 .60 - .69 Certain
7 .70 - .79 Certain
8 .80 - .89 Certain
9 .90 - .99 Certain
10 1.00 Certain
Fairly certain
Fairly uncertain





		Number Of Cells	Pct Of Map	Pct Of Area	Cumm Pct
0		2237	89.48	---	---
3	.30 - .39 Certain	15	0.60	5.70	5.70
4	.40 - .49 Certain	15	0.60	5.70	11.40
5	.50 - .59 Certain	71	2.84	27.00	38.40
6	.60 - .69 Certain	33	1.32	12.55	50.95
7	.70 - .79 Certain	118	4.72	44.87	95.82
8	.80 - .89 Certain	2	0.08	.76	96.58
9	.90 - .99 Certain	9	0.36	3.42	100.00
	Total	2500	100.00	100.00	---

Several comments and **guidelines on the recognition and assessment of error** in GIS processing have been promoted in papers on the subject. These are summarized below :

- There is a need for developing error statements for data contained within geographic information systems (Vitek et al, 1984).;
- The integration of data from different sources and in different original formats (e.g. points, lines, and areas), at different original scales, and possessing inherent errors can yield a product of questionable accuracy (Vitek et al, 1984).;
- The accuracy of a GIS-derived product is dependent on characteristics inherent in the source products, and on user requirements, such as scale of the desired output products and the method and resolution of data encoding (Marble, Peuquet, 1983).;
- The highest accuracy of any GIS output product can only be as accurate as the least accurate data theme of information involved in the analysis (Newcomer, Szajgin, 1984).;
- Accuracy of the data decreases as spatial resolution becomes more coarse (Walsh et al, 1987). ; and
- As the number of layers in an analysis increases, the number of possible opportunities for error increases (Newcomer, Szajgin, 1984).

2.0 Fundamental GIS Concepts

This chapter reviews the structural components and design of GIS data models. The focus is on reviewing spatial and attribute data models, and how data is encoded by the GIS software. This chapter describes GIS technical components and will be of most interest to technical staff and GIS operators.

2.1 Components of a GIS

The definition of *what exactly is a GIS ?* is a much discussed and debated question. A wide diversity of definitions and interpretations exist.

Overall, GIS should be viewed as a technology, not simply as a computer system.[8]

We commonly think of a GIS as a single, well-defined, integrated computer system. However, this is not always the case. A GIS can be made up of a variety of software and hardware tools. The important factor is the level of **integration** of these tools to provide a smoothly operating, fully functional geographic data processing environment.

In general, a GIS provides facilities for data capture, data management, data manipulation and analysis, and the presentation of results in both graphic and report form, with a particular emphasis upon preserving and utilizing inherent characteristics of spatial data.

The ability to incorporate spatial data, manage it, analyze it, and answer spatial questions is the distinctive characteristic of geographic information systems.[9]

A geographic information system, commonly referred to as a GIS, is an integrated set of hardware and software tools used for the manipulation and management of digital spatial (geographic) and related attribute data.

---

[8] Parker, 1988.

[9] Anderson, Starr, 1984.

A GIS has **four** main functional components.[10] These are :

- a data input subsystem;
- a data storage and retrieval subsystem;
- a data manipulation and analysis subsystem; and
- a data output and display subsystem.

A **data input** subsystem allows the user to capture, collect, and transform spatial and thematic data into digital form. The data inputs are usually derived from a combination of hard copy maps, aerial photographs, remotely sensed images, reports, survey documents, etc.

The **data storage and retrieval** subsystem organizes the data, spatial and attribute, in a form which permits it to be quickly retrieved by the user for analysis, and permits rapid and accurate updates to be made to the database. This component usually involves use of a database management system (DBMS) for maintaining attribute data. Spatial data is usually encoded and maintained in a proprietary file format.

The **data manipulation and analysis** subsystem allows the user to define and execute spatial and attribute procedures to generate derived information. This subsystem is commonly thought of as the *heart of a GIS*, and usually distinguishes it from other database information systems and computer-aided drafting (CAD) systems.

The **data output** subsystem allows the user to generate graphic displays, normally maps, and tabular reports representing derived information products.

The critical function for a GIS is, by design, the analysis of spatial data.[11]

It is important to understand that the GIS is not a new invention. In fact, geographic information processing has a rich history in a variety of disciplines. In particular, natural resource specialists and environmental scientists have been actively *processing* geographic data and promoting their techniques since the 1960's.

Today's generic, geographic information system, is distinguished from the geo-processing of the past by the use of computer automation to integrate geographic data processing tools in a friendly and comprehensive environment.

[10] Marble et al, 1984.

[11] Anderson, Starr, 1984.

The advent of sophisticated computer techniques has proliferated the multi-disciplinary application of geo-processing methodologies, and provided data integration capabilities that were logistically impossible before.

2.1.1 GIS Data Types

The basic premise of data in a GIS reflects traditional data found on a map. Accordingly, GIS technology utilizes two basic types of data. These are :

- spatial data (typically geographic); and
- attribute data.

Spatial data describes the absolute and relative location of geographic features.

Attribute data describes characteristics of the spatial features. These characteristics can be quantitative and/or qualitative in nature.

The coordinate location of a forestry stand would be spatial data, while the characteristics of that forestry stand, e.g. cover group, dominant species, crown closure, height, etc., would be attribute data. Other data types, in particular image and multimedia data, are becoming more prevalent with changing technology. Depending on the specific content of the data, *image data* may be considered either spatial, e.g. photographs, animation, movies, etc., or attribute, e.g. sound, descriptions, narration's, etc.

2.1.2 Physical Components of a GIS

Geographic information systems have **three** important physical components.[12] These are :

- Computer hardware;
- Computer software; and
- An organizational context.

All three components need to be balanced if the entire system is to function satisfactorily. The major hardware components of a GIS normally include the following :

- Central processing unit (CPU);
- Disk drives;
- Media drives, eg. tape, diskette, CD.;
- Digitizer;

[12] Burrough, 1986.

- Plotter (for maps and graphics);
- Printer (for reports_;
- Pointing device (mouse); and a
- Graphics display terminal.

The major software components of a GIS should satisfy the four subsystems identified earlier as well as be fully integrated with the relevant hardware of the system. Each of these subsystems is discussed in detail later on in the book.

The third physical component that is vital to the successful operation of a GIS is the **organizational context** of the company or agency that has purchased a GIS.

It is simply not sufficient for an organization to purchase a computer with some GIS software, hire some enthusiastic individuals and expect instant success.

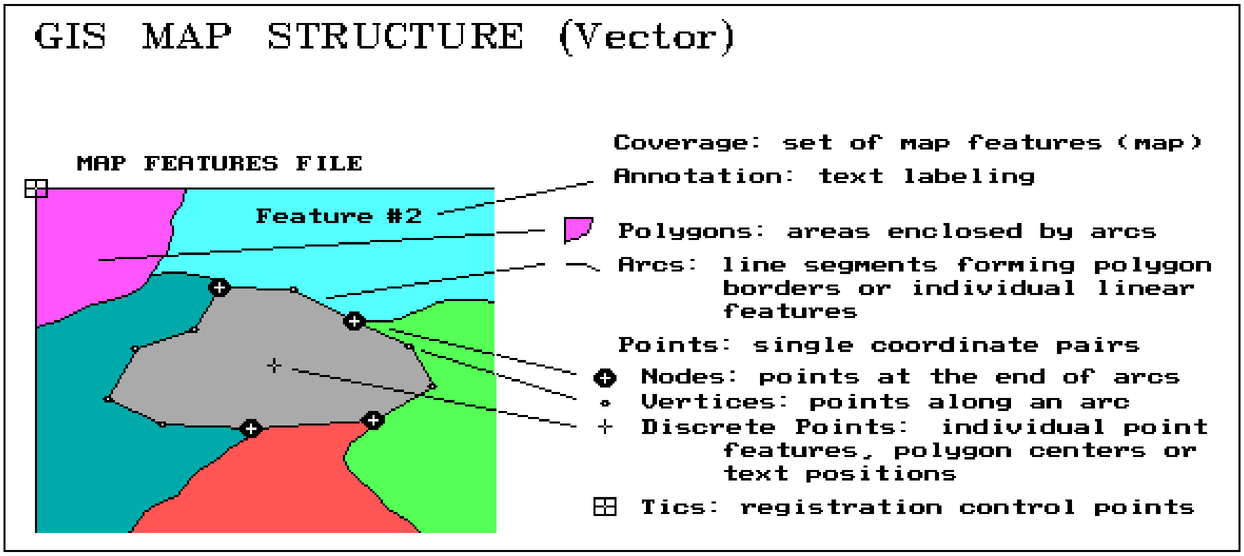
As in all organizations dealing with sophisticated technology, new tools can only be used effectively if they are properly integrated into the entire business strategy and operation. To do this properly requires not only the necessary investments in hardware and software, but also in the retraining and/or hiring of personnel to utilize the new technology in the proper organizational context.[13] Failure to implement your GIS without regard for a proper organizational commitment will result in an unsuccessful system ! Many of the issues concerned with organizational commitment are described in Chapter 8.

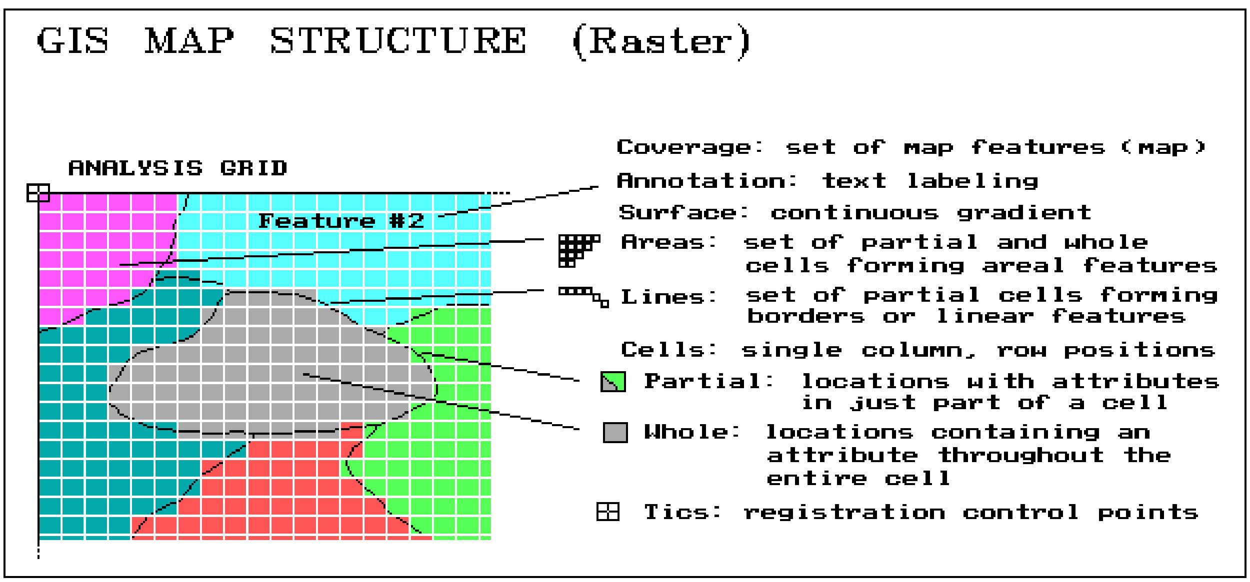
2.2 Spatial Data Models

Traditionally spatial data has been stored and presented in the form of a map. Three basic types of spatial data models have evolved for storing geographic data digitally. These are referred to as :

- **Vector**;
- **Raster;** and
- **Image.**

[13] Burrough, 1986.


GIS MAP STRUCTURE (Vector)
MAP FEATURES FILE
Feature #2
Coverage: set of map features (map)
Annotation: text labeling
Polygons: areas enclosed by arcs
Arcs: line segments forming polygon borders or individual linear features
Points: single coordinate pairs
Nodes: points at the end of arcs
Vertices: points along an arc
Discrete Points: individual point features, polygon centers or text positions
Tics: registration control points


GIS MAP STRUCTURE (Raster)
ANALYSIS GRID
Feature #2
Coverage: set of map features (map)
Annotation: text labeling
Surface: continuous gradient
Areas: set of partial and whole cells forming areal features
Lines: set of partial cells forming borders or linear features
Cells: single column, row positions
Partial: locations with attributes in just part of a cell
Whole: locations containing an attribute throughout the entire cell
Tics: registration control points

2.2.1 Vector Data Models

All spatial data models are approaches for storing the spatial location of geographic features in a database. Vector storage implies the use of vectors (directional lines) to represent a geographic feature. Vector data is characterized by the use of sequential points or *vertices* to define a linear segment. Each vertex consists of an X coordinate and a Y coordinate.

Vector lines are often referred to as *arcs* and consist of a string of vertices terminated by a *node*. A node is defined as a vertex that starts or ends an arc segment. Point features are defined by one coordinate pair, a vertex. Polygonal features are defined by a set of closed coordinate pairs. In vector representation, the storage of the vertices for each feature is important, as well as the connectivity between features, e.g. the sharing of common vertices where features connect.

Several different vector data models exist, however only two are commonly used in GIS data storage.

The most popular method of retaining spatial relationships among features is to explicitly record adjacency information in what is known as the *topologic data model*.[14] Topology is a mathematical concept that has its basis in the principles of feature adjacency and connectivity.

The topologic data structure is often referred to as an *intelligent data structure* because spatial relationships between geographic features are easily derived when using them. Primarily for this reason the topologic model is the dominant vector data structure currently used in GIS technology. Many of the complex data analysis functions cannot effectively be undertaken without a topologic vector data structure. Topology is reviewed in greater detail later on in the book.

The secondary vector data structure that is common among GIS software is the computer-aided drafting (**CAD)** data structure. This structure consists of listing elements, not features, defined by strings of vertices, to define geographic features, e.g. points, lines, or areas. There is considerable redundancy with this data model since the boundary segment between two polygons can be stored twice, once for each feature. The CAD structure emerged from the development of computer graphics systems without specific considerations of processing geographic features. Accordingly, since features, e.g. polygons, are self-contained and independent, questions about the adjacency of features can be difficult to answer. The CAD vector model lacks the definition of spatial relationships between features that is defined by the topologic data model.

[14] An excellent review of spatial data models is provided by Peuquet, 1984.

2.2.2 Raster Data Models

Raster data models incorporate the use of a *grid-cell* data structure where the geographic area is divided into cells identified by row and column. This data structure is commonly called *raster*. Whle the term raster implies a regularly spaced grid other *tessellated* data structures do exist in grid based GIS systems.[15] In particular, the quadtree data structure has found some acceptance as an alternative raster data model.

The size of cells in a tessellated data structure is selected on the basis of the data accuracy and the resolution needed by the user. There is no explicit coding of geographic coordinates required since that is implicit in the layout of the cells. A raster data structure is in fact a matrix where any coordinate can be quickly calculated if the origin point is known, and the size of the grid cells is known. Since grid-cells can be handled as two-dimensional arrays in computer encoding many analytical operations are easy to program. This makes tessellated data structures a popular choice for many GIS software. Topology is not a relevant concept with tessellated structures since adjacency and connectivity are implicit in the location of a particular cell in the data matrix.

Several tessellated data structures exist, however only **two** are commonly used in GIS's. The most popular cell structure is the regularly spaced matrix or *raster* structure. This data structure involves a division of spatial data into regularly spaced cells. Each cell is of the same shape and size. Squares are most commonly utilized.

Since geographic data is rarely distinguished by regularly spaced shapes, cells must be classified as to the most common attribute for the cell. The problem of determining the proper resolution for a particular data layer can be a concern. If one selects too coarse a cell size then data may be overly generalized. If one selects too fine a cell size then too many cells may be created resulting in a large data volumes, slower processing times, and a more cumbersome data set. As well, one can imply an accuracy greater than that of the original data capture process and this may result in some erroneous results during analysis.

As well, since most data is captured in a vector format, e.g. digitizing, data must be converted to the raster data structure. This is called *vector-raster conversion.* Most GIS software allows the user to define the raster grid (cell) size for vector-raster conversion. It is imperative that the original scale, e.g. accuracy, of the data be known prior to conversion. The accuracy of the data, often referred to as the resolution, should determine the cell size of the output raster map during conversion.

Most raster based GIS software requires that the raster cell contain only a single discrete value. Accordingly, a data layer, e.g. forest inventory stands, may be broken down into a series of raster maps, each representing an attribute type, e.g. a species map, a height map, a density map, etc. These are often referred to as *one attribute maps*. This is in contrast to most conventional vector data models that maintain data as *multiple attribute maps*, e.g. forest inventory polygons *linked* to a database table containing all attributes as columns. This basic distinction of raster data storage provides the foundation for quantitative analysis techniques. This is often referred to as *raster or map algebra*. The use of raster data structures allow for

[15] Peuquet, 1984.

sophisticated mathematical modelling processes while vector based systems are often constrained by the capabilities and language of a relational DBMS.

This difference is the major distinguishing factor between vector and raster based GIS software. It is also important to understand that the selection of a particular data structure can provide advantages during the analysis stage. For example, the vector data model does not handle continuous data, e.g. elevation, very well while the raster data model is more ideally suited for this type of analysis. Accordingly, the raster structure does not handle linear data analysis, e.g. shortest path, very well while vector systems do. It is important for the user to understand that there are certain advantages and disadvantages to each data model.

The selection of a particular data model, vector or raster, is dependent on the source and type of data, as well as the intended use of the data. Certain analytical procedures require raster data while others are better suited to vector data.

2.2.3 Image Data

Image data is most often used to represent graphic or pictorial data. The term *image* inherently reflects a graphic representation, and in the GIS world, differs significantly from raster data. Most often, image data is used to store remotely sensed imagery, e.g. satellite scenes or orthophotos, or ancillary graphics such as photographs, scanned plan documents, etc. Image data is typically used in GIS systems as background display data (if the image has been rectified and georeferenced); or as a graphic attribute. Remote sensing software makes use of image data for image classification and processing. Typically, this data must be converted into a raster format (and perhaps vector) to be used analytically with the GIS.

Image data is typically stored in a variety of de facto industry standard proprietary formats. These often reflect the most popular image processing systems. Other graphic image formats, such as TIFF, GIF, PCX, etc., are used to store ancillary image data. Most GIS software will read such formats and allow you to display this data.

2.2.4 Advantages and Disadvantages

There are several advantages and disadvantages for using either the vector or raster data model to store spatial data.

Vector Data Model

Advantages :

- Data can be represented at its original resolution and form without generalization.
- Graphic output is usually more aesthetically pleasing (traditional cartographic representation);
- Since most data, e.g. hard copy maps, is in vector form no data conversion is required.

- Accurate geographic location of data is maintained.
- Allows for efficient encoding of topology, and as a result more efficient operations that require topological information, e.g. proximity, network analysis.

Disadvantages :

- The location of each vertex needs to be stored explicitly.
- For effective analysis, vector data must be converted into a topological structure. This is often processing intensive and usually requires extensive data cleaning. As well, topology is static, and any updating or editing of the vector data requires re-building of the topology.;
- Algorithms for manipulative and analysis functions are complex and may be processing intensive. Often, this inherently limits the functionality for large data sets, e.g. a large number of features.;
- Continuous data, such as elevation data, is not effectively represented in vector form. Usually substantial data generalization or interpolation is required for these data layers.; and
- Spatial analysis and filtering within polygons is impossible.

Raster Data Model

Advantages :

- The geographic location of each cell is implied by its position in the cell matrix. Accordingly, other than an origin point, e.g. bottom left corner, no geographic coordinates are stored.;
- Due to the nature of the data storage technique data analysis is usually easy to program and quick to perform.;
- The inherent nature of raster maps, e.g. one attribute maps, is ideally suited for mathematical modeling and quantitative analysis.;
- Discrete data, e.g. forestry stands, is accommodated equally well as continuous data, e.g. elevation data, and facilitates the integrating of the two data types.; and
- Grid-cell systems are very compatible with raster-based output devices, e.g. electrostatic plotters, graphic terminals.

Disadvantages :

- The cell size determines the resolution at which the data is represented.;
- It is especially difficult to adequately represent linear features depending on the cell resolution. Accordingly, network linkages are difficult to establish.;
- Processing of associated attribute data may be cumbersome if large amounts of data exists. Raster maps inherently reflect only one attribute or characteristic for an area.;
- Since most input data is in vector form, data must undergo vector-to-raster conversion. Besides increased processing requirements this may introduce data integrity concerns due to generalization.;
- Most output maps from grid-cell systems do not conform to high-quality graphic needs.

It is often difficult to compare or rate GIS software that use different data models. Some personal computer (PC) packages utilize vector structures for data input, editing, and display but convert to raster structures for any analysis. Other more comprehensive GIS offerings provide both integrated raster and vector analysis techniques. They allow users to select the data structure appropriate for the analysis requirements. Integrated raster and vector processing capabilities are most desirable and provide the greatest flexibility for data manipulation and analysis.

2.3 Attribute Data Models

A separate data model is used to store and maintain attribute data for GIS software. These data models may exist internally within the GIS software, or may be reflected in external commercial Database Management Software (DBMS). A variety of different data models exist for the storage and management of attribute data. The most common are :

- Tabular
- Hierarchial
- Network
- Relational
- Object Oriented

The tabular model is the manner in which most early GIS software packages stored their attribute data. The next three models are those most commonly implemented in database management systems (DBMS). The object oriented is newer but rapidly gaining in popularity for some applications.[16] A brief review of each model is provided.

2.3.1 Tabular Model

The simple tabular model stores attribute data as sequential data files with fixed formats (or comma delimited for ASCII data), for the location of attribute values in a predefined record structure. This type of data model is outdated in the GIS arena. It lacks any method of checking data integrity, as well as being inefficient with respect to data storage, e.g. limited indexing capability for attributes or records, etc.

2.3.2 Hierarchial Model

The hierarchial database organizes data in a *tree* structure. Data is structured downward in a *hierarchy* of tables. Any level in the hierarchy can have unlimited *children*, but any *child* can have only one *parent*. **Hierarchial DBMS have not gained any noticeable acceptance for use within GIS**. They are oriented for data sets that are very stable, where primary relationships among the data change infrequently or never at all. Also, the limitation on the

[16] Aronson, 1987.

number of parents that an element may have is not always conducive to actual geographic phenomenon.

2.3.3 Network Model

The network database organizes data in a network or *plex* structure. Any column in a plex structure can be linked to any other. Like a tree structure, a plex structure can be described in terms of *parents* and *children*. This model allows for children to have more than one parent.

Network DBMS have not found much more acceptance in GIS than the hierarchial DBMS. They have the same flexibility limitations as hierarchial databases; however, the more powerful structure for representing data relationships allows a more realistic modelling of geographic phenomenon. However, network databases tend to become overly complex too easily. In this regard it is easy to lose control and understanding of the relationships between elements.

2.3.4 Relational Model

The relational database organizes data in *tables*. Each table, is identified by a unique table name, and is organized by *rows* and *columns*. Each column within a table also has a unique name. Columns store the values for a specific attribute, e.g. cover group, tree height. Rows represent one record in the table. In a GIS each row is usually linked to a separate spatial feature, e.g. a forestry stand. Accordingly, each row would be comprised of several columns, each column containing a specific value for that geographic feature. The following figure presents a sample table for forest inventory features. This table has 4 rows and 5 columns. The forest stand number would be the *label* for the spatial feature as well as the *primary key* for the database table. This serves as the linkage between the spatial definition of the feature and the attribute data for the feature.

STAND NUMBER	COVER GROUP	HEIGHT	SITE INDEX	STAND AGE
001	DEC	3	G	100
002	DEC-CON	4	M	80
003	DEC-CON	4	M	60
004	CON	4	G	120

Data is often stored in several tables. Tables can be joined or referenced to each other by common columns (relational fields). Usually the common column is an identification number for a selected geographic feature, e.g. a forestry stand polygon number. This identification number acts as the *primary key* for the table. The ability to join tables through use of a common column is the essence of the relational model. Such relational joins are usually ad hoc in nature and form the basis of for querying in a relational GIS product. Unlike the other previously discussed database types, relationships are implicit in the character of the data as opposed to explicit characteristics of the database set up.[17]

The relational database model is the most widely accepted for managing the attributes of geographic data.

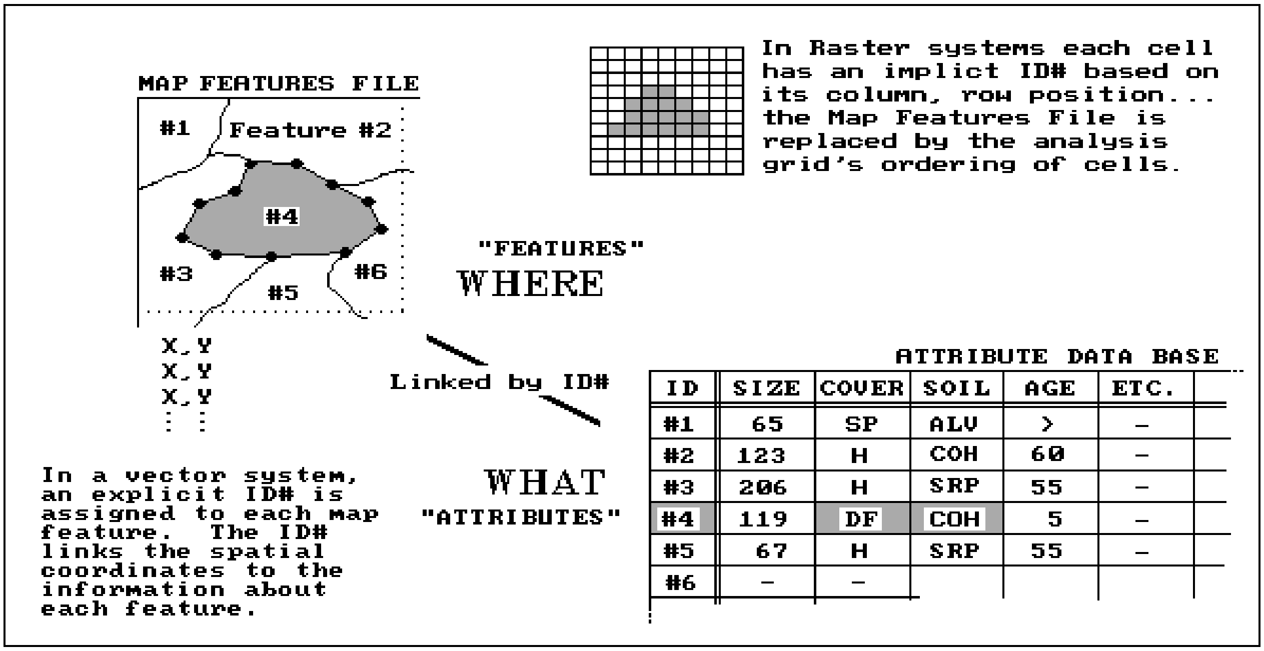
In fact, most GIS software provides an *internal* relational data model, as well as support for *commercial off-the-shelf* (COTS) relational DBMS'. COTS DBMS' are referred to as *external* DBMS'. This approach supports both users with small data sets, where an internal data model is sufficient, and customers with larger data sets who utilize a DBMS for other corporate data storage requirements. With an external DBMS the GIS software can simply *connect* to the database, and the user can make use of the inherent capabilities of the DBMS. External DBMS' tend to have much more extensive querying and data integrity capabilities than the GIS' internal relational model. The emergence and use of the external DBMS is a trend that has resulted in the proliferation of GIS technology into more traditional data processing environments.

The relational DBMS is attractive because of its :

- simplicity in organization and data modelling.;
- flexibility - data can be manipulated in an ad hoc manner by joining tables.;
- efficiency of storage - by the proper design of data tables redundant data can be minimized; and
- the non-procedural nature - queries on a relational database do not need to take into account the internal organization of the data.

The relational DBMS has emerged as the dominant commercial data management tool in GIS implementation and application.

[17] Aronson, 1987.



This figure illustrates the basic linkage between a vector spatial data (topologic model) and attributes maintained in a relational database file.

2.3.5 Object-Oriented Model

The object-oriented database model manages data through *objects*. An object is a collection of data elements and operations that together are considered a single entity. The object-oriented database is a relatively new model. This approach has the attraction that querying is very natural, as features can be bundled together with attributes at the database administrator's discretion.[18] To date, only a few GIS packages are promoting the use of this attribute data model. However, initial impressions indicate that this approach may hold many operational benefits with respect to geographic data processing. Fulfilment of this promise with a commercial GIS product remains to be seen.19

2.4 Spatial Data Relationships

The nature of spatial data relationships are important to understand within the context of GIS. In particular, the relationship between geographic features is a complex problem in which we are far from understanding in its entirety. This is of concern since the primary role of GIS is the manipulation and analysis of large quantities of spatial data. To date, the accepted theoretical solution is to *topologically structure* spatial data.

---

18 Aronson, 1987.

19 An excellent overview of Object Oriented concepts can be found in a Vol. 14, No. 3 - Fall 1996 issue of The Compiler (see References for contact address). The article is entitled "Object-Oriented Technology and Its GIS Expressions, Dr. Joseph K. Berry. Other more detailed descriptions can be found in the 'Beyond Mapping' column of GIS World, vol.6:12, 7-1, 8:1-3, 8:12, and 9:1-2.

It is believed that a topologic data model best reflects the geography of the real world and provides an effective mathematical foundation for encoding spatial relationships, providing a data model for manipulating and analyzing vector based data.

Most GIS software segregate spatial and attribute data into separate data management systems. Most frequently, the topological or raster structure is used to store the spatial data, while the relational database structure is used to store the attribute data. Data from both structures are linked together for use through unique identification numbers, e.g. feature labels and DBMS primary keys. This coupling of spatial features with an attribute record is usually maintained by an internal number assigned by the GIS software. A *label* is required so the user can load the appropriate attribute record for a given geographic feature. Most often a single attribute record is automatically created by the GIS software once a clean topological structure is properly generated. This attribute record normally contains the internal number for the feature, the user's label identifier, the area of the feature, and the perimeter of the feature. Linear features have the length of the feature defined instead of the area.

2.4.1 Topology

The topologic model is often confusing to initial users of GIS. Topology is a mathematical approach that allows us to structure data based on the principles of feature adjacency and feature connectivity. It is in fact the mathematical method used to define spatial relationships. Without a topologic data structure in a vector based GIS most data manipulation and analysis functions would not be practical or feasible.

The most common topological data structure is the *arc/node* data model. This model contains two basic entities, the *arc* and the *node*. The arc is a series of points, joined by straight line segments, that start and end at a node. The node is an intersection point where two or more arcs meet. Nodes also occur at the end of a *dangling* arc, e.g. an arc that does not connect to another arc such as a dead end street. Isolated nodes, not connected to arcs represent point features. A polygon feature is comprised of a closed chain of arcs.

In GIS software the topological definition is commonly stored in a proprietary format. However, most software offerings record the topological definition in three tables. These tables are analogous to relational tables. The three tables represent the different types of features, e.g. point, line, area. A fourth table containing the coordinates is also utilized. The *node table* stores information about the node and the arcs that are connected to it. The *arc table* contains topological information about the arcs. This includes the start and end node, and the polygon to the left and right that the arc is an element of. The *polygon table* defines the arcs that make up each polygon. While arc, node, and polygon terminology is used by most GIS vendors, some also introduce terms such as *edges* and *faces* to define arcs and polygons. This is merely the use of different words to define topological definitions. Do not be confused by this.

Since most input data does not exist in a topological data structure, topology must be *built* with the GIS software. Depending on the data set this can be an CPU intensive and time consuming procedure. This building process involves the creation of the topological tables and the definition of the arc, node, and polygon entities. To properly define the topology there are specific requirements with respect to graphic elements, e.g. no duplicate lines, no gaps in arcs that define polygon features, etc. These requirements are reviewed in the Data Editing section of the book.

The topological model is utilized because it effectively models the relationship of spatial entities. Accordingly, it is well suited for operations such as contiguity and connectivity analyses. Contiguity involves the evaluation of feature adjacency, e.g. features that touch one another, and proximity, e.g. features that are near one another. The primary advantage of the topological model is that spatial analysis can be done without using the coordinate data. Many operations can be done largely, if not entirely, by using the topological definition alone. This is a significant advantage over the CAD or *spaghetti* vector data structure that requires the derivation of spatial relationships from the coordinate data before analysis can be undertaken.

The major disadvantage of the topological data model is its static nature. It can be a time consuming process to properly define the topology depending on the size and complexity of the data set. For example, 2,000 forest stand polygons will require considerably longer to *build* the topology that 2,000 municipal lot boundaries. This is due to the inherent complexity of the features, e.g. lots tend to be rectangular while forest stands are often long and sinuous. This can be a consideration when evaluating the topological building capabilities of GIS software. The static nature of the topological model also implies that every time some editing has occurred, e.g. forest stand boundaries are changed to reflect harvesting or burns, the topology must be rebuilt. The integrity of the topological structure and the DBMS tables containing the attribute data can be a concern here.[20] This is often referred to as *referential integrity*. While topology is the mechanism to ensure integrity with spatial data, referential integrity is the concept of ensuring integrity for both linked topological data and attribute data.

[20] Many of the issues raised with the topologic model are described in detail by Aronoff, 1989.

3.0 Sources of Data and Data Input Techniques

This chapter reviews different sources, formats, and input techniques for GIS data. The focus is on reviewing different data input techniques for spatial data. This chapter also describes data input errors, spatial and attribute, and reviews typical procedures to correct input errors. This chapter will be of most interest to technical staff and GIS operators.

3.1 Sources of Data

As previously identified, two types of data are input into a GIS, spatial and attribute. The data input process is the operation of encoding both types of data into the GIS database formats.

The creation of a clean digital database is the most important and time consuming task upon which the usefulness of the GIS depends.[21] The establishment of a robust spatial database is the cornerstone of a successful GIS implementation.

As well, the digital data is the most expensive part of the GIS. Yet, often, not enough attention is given to the quality of the data or the processes by which they are prepared for automation.[22]

The general consensus among the GIS community is that 60 to 80 % of the cost incurred during implementation of GIS technology lies in data acquisition, data compilation and database development.

A wide variety of data sources exist for both spatial and attribute data. The most common general sources for spatial data are :

- hard copy maps;
- aerial photographs;
- remotely-sensed imagery;
- point data samples from surveys; and
- existing digital data files.

---

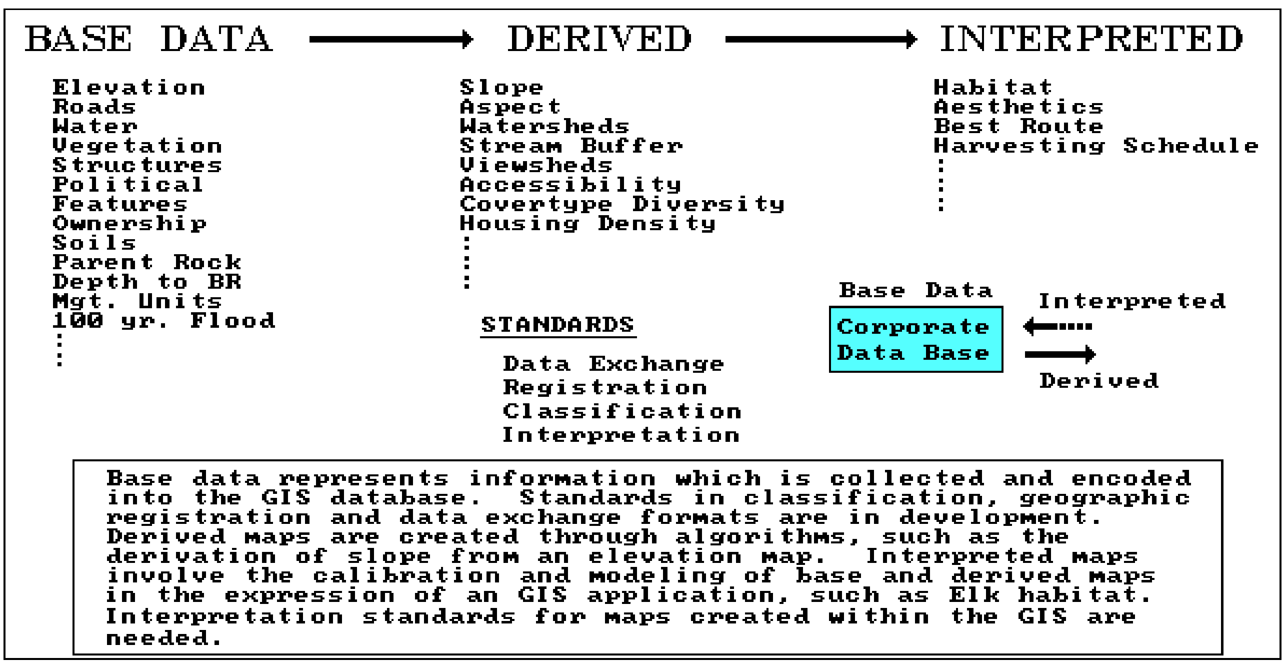
[21] Burrough, 1986.

[22] Dangermond, 1987.

Existing hard copy maps, e.g. sometimes referred to as *analogue maps*, provide the most popular source for any GIS project.

Potential users should be aware that while there are many private sector firms specializing in providing digital data, federal, provincial and state government agencies are an excellent source of data. Because of the large costs associated with data capture and input, government departments are often the only agencies with financial resources and manpower funding to invest in data compilation. British Columbia and Alberta government agencies are good examples. Both provincial governments have defined and implemented province wide coverage of digital base map data at varying map scales, e.g. 1:20,000 and 1:250,000. As well, the provincial forestry agencies also provide thematic forest inventory data in digital format. Federal agencies are also often a good source for base map information. An inherent advantage of digital data from government agencies is its cost. It is typically inexpensive. However, this is often offset by the data's accuracy and quality. Thematic coverages are often not up to date. However, it is important to note that specific characteristics of government data varies greatly across North America.

Attribute data has an even wider variety of data sources. Any textual or tabular data than can be referenced to a geographic feature, e.g. a point, line, or area, can be input into a GIS. Attribute data is usually input by manual keying or via a bulk loading utility of the DBMS software. ASCII format is a de facto standard for the transfer and conversion of attribute information.



This figure describes the basic data types that are used and created by a GIS.

3.2 Data Input Techniques

Since the input of attribute data is usually quite simple, the discussion of data input techniques will be limited to spatial data only. There is no single method of entering the spatial data into a GIS. Rather, there are several, mutually compatible methods that can be used singly or in combination.

The choice of data input method is governed largely by the application, the available budget, and the type and the complexity of data being input.[23]

There are at least four basic procedures for inputting spatial data into a GIS.[24] These are :

- Manual digitizing;
- Automatic scanning;
- Entry of coordinates using coordinate geometry; and the
- Conversion of existing digital data.

3.2.1 Digitizing

While considerable work has been done with newer technologies, the overwhelming majority of GIS spatial data entry is done by manual digitizing.[25] A digitizer is an electronic device consisting of a table upon which the map or drawing is placed. The user traces the spatial features with a hand-held magnetic pen, often called a *mouse* or cursor. While tracing the features the coordinates of selected points, e.g. vertices, are sent to the computer and stored. All points that are recorded are registered against positional control points, usually the map corners, that are keyed in by the user at the beginning of the digitizing session. The coordinates are recorded in a user defined coordinate system or map projection. Latitude and longitude and UTM is most often used. The ability to adjust or transform data during digitizing from one projection to another is a desirable function of the GIS software. Numerous functional techniques exist to aid the operator in the digitizing process.

Digitizing can be done in a *point mode*, where single points are recorded one at a time, or in a *stream mode*, where a point is collected on regular intervals of time or distance, measured by an X and Y movement, e.g. every 3 metres. Digitizing can also be done blindly or with a graphics terminal. Blind digitizing infers that the graphic result is not immediately viewable to the person digitizing. Most systems display the digitized linework as it is being digitized on an accompanying graphics terminal.

Most GIS's use a *spaghetti mode* of digitizing. This allows the user to simply digitize

[23] Burrough, 1986.

[24] Dangermond, 1987.

[25] Dangermond, 1987.

lines by indicating a start point and an end point. Data can be captured in point or stream mode. However, some systems do allow the user to capture the data in an arc/node topological data structure. The arc/node data structure requires that the digitizer identify nodes.

Data capture in an arc/node approach helps to build a topologic data structure immediately. This lessens the amount of post processing required to *clean* and build the topological definitions. However, most often digitizing with an arc/node approach does not negate the requirement for editing and cleaning of the digitized linework before a complete topological structure can be obtained.

The building of topology is primarily a post-digitizing process that is commonly executed in *batch mode* after data has been cleaned. To date, only a few commercial vector GIS software offerings have successfully exhibited the capability to build topology interactively while the user digitizes.

Manual digitizing has many advantages. These include :

- Low capital cost, e.g. digitizing tables are cheap;
- Low cost of labour;
- Flexibility and adaptability to different data types and sources;
- Easily taught in a short amount of time - an easily mastered skill;
- Generally the quality of data is high;
- Digitizing devices are very reliable and most often offer a greater precision that the data warrants; and
- Ability to easily register and update existing data.

For raster based GIS software data is still commonly digitized in a vector format and converted to a raster structure after the building of a clean topological structure. The procedure usually differs minimally from vector based software digitizing, other than some raster systems allow the user to define the resolution size of the grid-cell. Conversion to the raster structure may occur *on-the-fly* or afterwards as a separate conversion process.

3.2.2 Automatic Scanning

A variety of scanning devices exist for the automatic capture of spatial data. While several different technical approaches exist in scanning technology, all have the advantage of being able to capture spatial features from a map at a rapid rate of speed. However, as of yet, scanning has not proven to be a viable alternative for most GIS implementation. Scanners are generally expensive to acquire and operate. As well, most scanning devices have limitations with respect to the capture of selected features, e.g. text and symbol recognition. Experience has shown that most scanned data requires a substantial amount of manual editing to create a clean data layer. Given these basic constraints some other practical limitations of scanners should be identified. These include :

- hard copy maps are often unable to be removed to where a scanning device is available, e.g. most companies or agencies cannot afford their own scanning device and therefore must send their maps to a private firm for scanning;

- hard copy data may not be in a form that is viable for effective scanning, e.g. maps are of poor quality, or are in poor condition;

- geographic features may be too few on a single map to make it practical, cost-justifiable, to scan;

- often on *busy* maps a scanner may be unable to distinguish the features to be captured from the surrounding graphic information, e.g. dense contours with labels;

- with raster scanning there it is difficult to read unique labels (text) for a geographic feature effectively; and

- scanning is much more expensive than manual digitizing, considering all the cost/performance issues.

Consensus within the GIS community indicates that scanners work best when the information on a map is kept very clean, very simple, and uncluttered with graphic symbology.

The sheer cost of scanning usually eliminates the possibility of using scanning methods for data capture in most GIS implementations. Large data capture shops and government agencies are those most likely to be using scanning technology.

Currently, general consensus is that the quality of data captured from scanning devices is not substantial enough to justify the cost of using scanning technology. However, major breakthroughs are being made in the field, with scanning techniques and with capabilities to automatically clean and prepare scanned data for topological encoding. These include a variety of *line following* and *text recognition* techniques. Users should be aware that this technology has great potential in the years to come, particularly for larger GIS installations.

3.2.3 Coordinate Geometry

A third technique for the input of spatial data involves the calculation and entry of coordinates using coordinate geometry (COGO) procedures. This involves entering, from survey data, the explicit measurement of features from some known monument. This input technique is obviously very costly and labour intensive. In fact, it is rarely used for natural resource applications in GIS. This method is useful for creating very precise cartographic definitions of property, and accordingly is more appropriate for land records management at the cadastral or municipal scale.

3.2.4 Conversion of Existing Digital Data

A fourth technique that is becoming increasingly popular for data input is the conversion of existing digital data. A variety of spatial data, including digital maps, are openly available from a wide range of government and private sources. The most common digital data to be used in a GIS is data from CAD systems. A number of data conversion programs exist, mostly from GIS software vendors, to transform data from CAD formats to a raster or topological GIS data format. Several ad hoc standards for data exchange have been established in the market place. These are supplemented by a number of government distribution formats that have been developed. Given the wide variety of data formats that exist, most GIS vendors have developed and provide data exchange/conversion software to go from their format to those considered common in the market place.

Most GIS software vendors also provide an ASCII data exchange format specific to their product, and a programming subroutine library that will allow users to write their own data conversion routines to fulfil their own specific needs. As digital data becomes more readily available this capability becomes a necessity for any GIS. Data conversion from existing digital data is not a problem for most technical persons in the GIS field. However, for smaller GIS installations who have limited access to a *GIS analyst* this can be a major stumbling block in getting a GIS operational. Government agencies are usually a good source for technical information on data conversion requirements.

Some of the data formats common to the GIS marketplace are listed below. Please note that most formats are only utilized for graphic data. Attribute data is usually handled as ASCII text files. Vendor names are supplied where appropriate.

IGDS - Interactive Graphics Design Software (Intergraph)

This binary format is a standard in the turnkey CAD market and has become a de facto standard in Canada's mapping industry. It is a proprietary format, however most GIS software vendors provide DGN translators.

DLG - Digital Line Graph (US Geological Survey)

This ASCII format is used by the USGS as a distribution standard and consequently is well utilized in the United States. It is not used very much in Canada even though most software vendors provide two way conversion to DLG.

DXF - Drawing Exchange Format (Autocad)

This ASCII format is used primarily to convert to/from the Autocad drawing format and is a standard in the engineering discipline. Most GIS software vendors provide a DXF translator.

GENERATE - ARC/INFO Graphic Exchange Format

A generic ASCII format for spatial data used by the ARC/INFO software to accommodate generic spatial data.

EXPORT - ARC/INFO Export Format

An exchange format that includes both graphic and attribute data. This format is intended for transferring ARC/INFO data from one hardware platform, or site, to another. It is also often used for archiving ARC/INFO data. This is not a published data format, however some GIS and desktop mapping vendors provide translators. EXPORT format can come in either uncompressed, partially compressed, or fully compressed format.

A wide variety of other vendor specific data formats exist within the mapping and GIS industry. In particular, most GIS software vendors have their own proprietary formats. However, almost all provide data conversion to/from the above formats. As well, most GIS software vendors will develop data conversion programs dependant on specific requests by customers. Potential purchasers of commercial GIS packages should determine and clearly identify their data conversion needs, prior to purchase, to the software vendor.

3.3 Data Editing

Data editing and verification is in response to the errors that arise during the encoding of spatial and non-spatial data. The editing of spatial data is a time consuming, interactive process that can take as long, if not longer, than the data input process itself.

Several kinds of errors can occur during data input.[26] They can be classified as :

- **Incompleteness of the spatial data**.This includes missing points, line segments, and/or polygons.
- **Locational placement errors of spatial data**. These types of errors usually are the result of careless digitizing or poor quality of the original data source.
- **Distortion of the spatial data**. This kind of error is usually caused by base maps that are not scale-correct over the whole image, e.g. aerial photographs, or from material stretch, e.g. paper documents.
- **Incorrect linkages between spatial and attribute data**. This type of error is commonly the result of incorrect unique identifiers (labels) being assigned during manual key in or digitizing. This may involve the assigning of an entirely wrong label to a feature, or more than one label being assigned to a feature.
- **Attribute data is wrong or incomplete**. Often the attribute data does not match exactly with the spatial data. This is because they are frequently from independent sources and often different time periods. Missing data records or too many data records are the most common problems.

The identification of errors in spatial and attribute data is often difficult. Most spatial errors become evident during the topological building process. The use of *check plots* to clearly determine where spatial errors exist is a common practice. Most topological building functions in GIS software clearly identify the geographic location of the error and indicate the nature of the

[26] after Burrough, 1986.

problem. Comprehensive GIS software allows users to graphically walk through and edit the spatial errors. Others merely identify the type and coordinates of the error. Since this is often a labour intensive and time consuming process, users should consider the error correction capabilities very important during the evaluation of GIS software offerings.

3.3.1 Spatial Data Errors

A variety of common data problems occur in converting data into a topological structure. These stem from the original quality of the source data and the characteristics of the data capture process. Usually data is input by digitizing. Digitizing allows a user to trace spatial data from a hard copy product, e.g. a map, and have it recorded by the computer software. Most GIS software has utilities to *clean* the data and build a topologic structure. If the data is unclean to start with, for whatever reason, the cleaning process can be very lengthy. Interactive editing of data is a distinct reality in the data input process.

Experience indicates that in the course of any GIS project 60 to 80 % of the time required to complete the project is involved in the input, cleaning, linking, and verification of the data.

The most common problems that occur in converting data into a topological structure include :

- slivers and gaps in the line work;
- dead ends, e.g. also called dangling arcs, resulting from overshoots and undershoots in the line work; and
- bow ties or weird polygons from inappropriate closing of connecting features.

Of course, topological errors only exist with linear and areal features. They become most evident with polygonal features. *Slivers* are the most common problem when cleaning data. Slivers frequently occur when coincident boundaries are digitized separately, e.g. once each for adjacent forest stands, once for a lake and once for the stand boundary, or after polygon overlay. Slivers often appear when combining data from different sources, e.g. forest inventory, soils, and hydrography. It is advisable to digitize data layers with respect to an existing data layer, e.g. hydrography, rather than attempting to match data layers later. A proper plan and definition of priorities for inputting data layers will save many hours of interactive editing and cleaning.

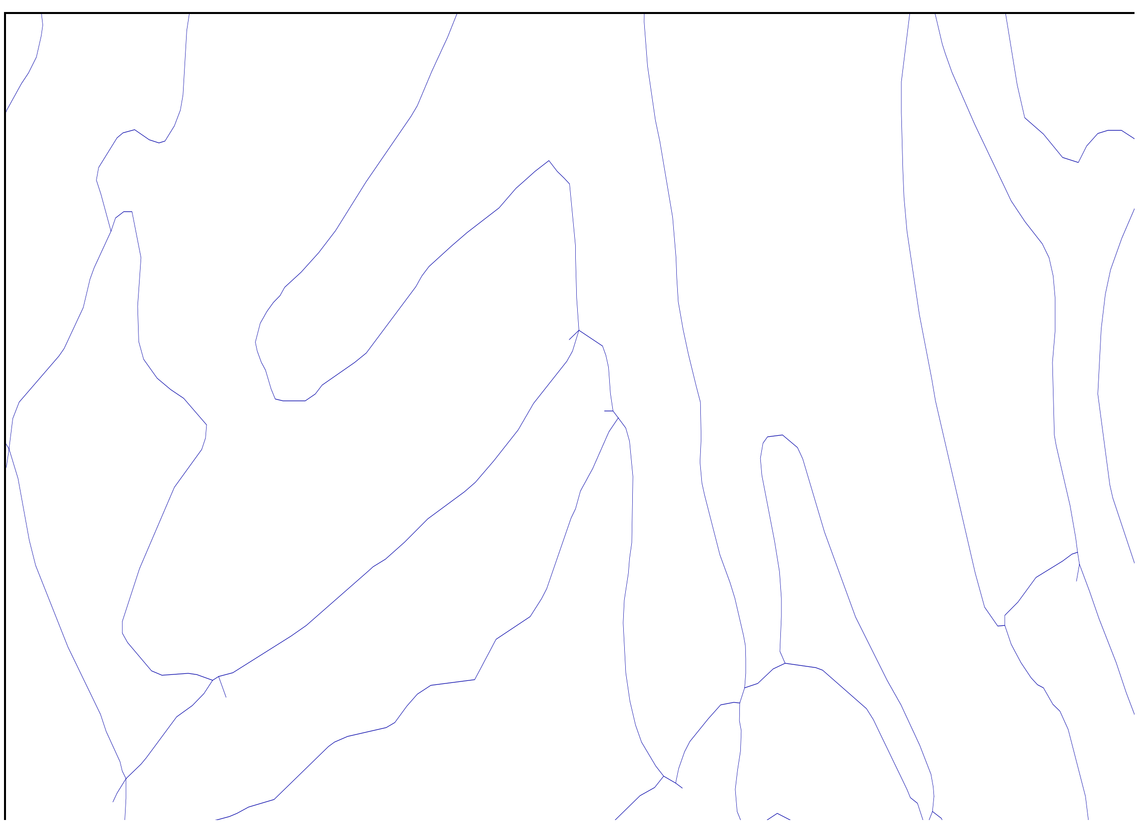
Dead ends usually occur when data has been digitized in a *spaghetti* mode, or without snapping to existing nodes. Most GIS software will clean up undershoots and overshoots based on a user defined tolerance, e.g. distance. The definition of an inappropriate distance often leads to the formation of *bow ties* or *weird polygons* during topological building. Tolerances that are too large will force arcs to snap one another that should not be connected. The result is small polygons called *bow ties*. The definition of a proper tolerance for cleaning requires an understanding of the scale and accuracy of the data set.

The other problem that commonly occurs when building a topologic data structure is *duplicate lines*. These usually occur when data has been digitized or converted from a CAD system. The lack of topology in these type of drafting systems permits the inadvertent creation of elements that are exactly duplicate. However, most GIS packages afford automatic elimination of duplicate elements during the topological building process. Accordingly, it may not be a concern with vector based GIS software. Users should be aware of the duplicate element that retraces itself, e.g. a three vertice line where the first point is also the last point. Some GIS packages do not identify these feature inconsistencies and will build such a feature as a valid polygon. This is because the topological definition is mathematically correct, however it is not geographically correct. Most GIS software will provide the capability to eliminate bow ties and slivers by means of a feature elimination command based on area, e.g. polygons less than 100 square metres. The ability to define custom topological error scenarios and provide for semi-automated correction is a desirable capability for GIS software.

The adjoining figure illustrates some typical errors described above. Can you spot them ? They include undershoots, overshoots, bow ties, and slivers. Most bow ties occur when inappropriate tolerances are used during the automated cleaning of data that contains many overshoots. This particular set of spatial data is a prime candidate for numerous bow tie polygons.

3.3.2 Attribute Data Errors

The identification of attribute data errors is usually not as simple as spatial errors. This is



especially true if these errors are attributed to the quality or reliability of the data. Errors as such usually do not surface until later on in the GIS processing. Solutions to these type of problems are much more complex and often do not exist entirely. It is much more difficult to spot errors in attribute data when the values are syntactically good, but incorrect.[27]

Simple errors of linkage, e.g. missing or duplicate records, become evident during the linking operation between spatial and attribute data. Again, most GIS software contains functions that check for and clearly identify problems of linkage during attempted operations. This is also an area of consideration when evaluating GIS software.

3.3.3 Data Verification

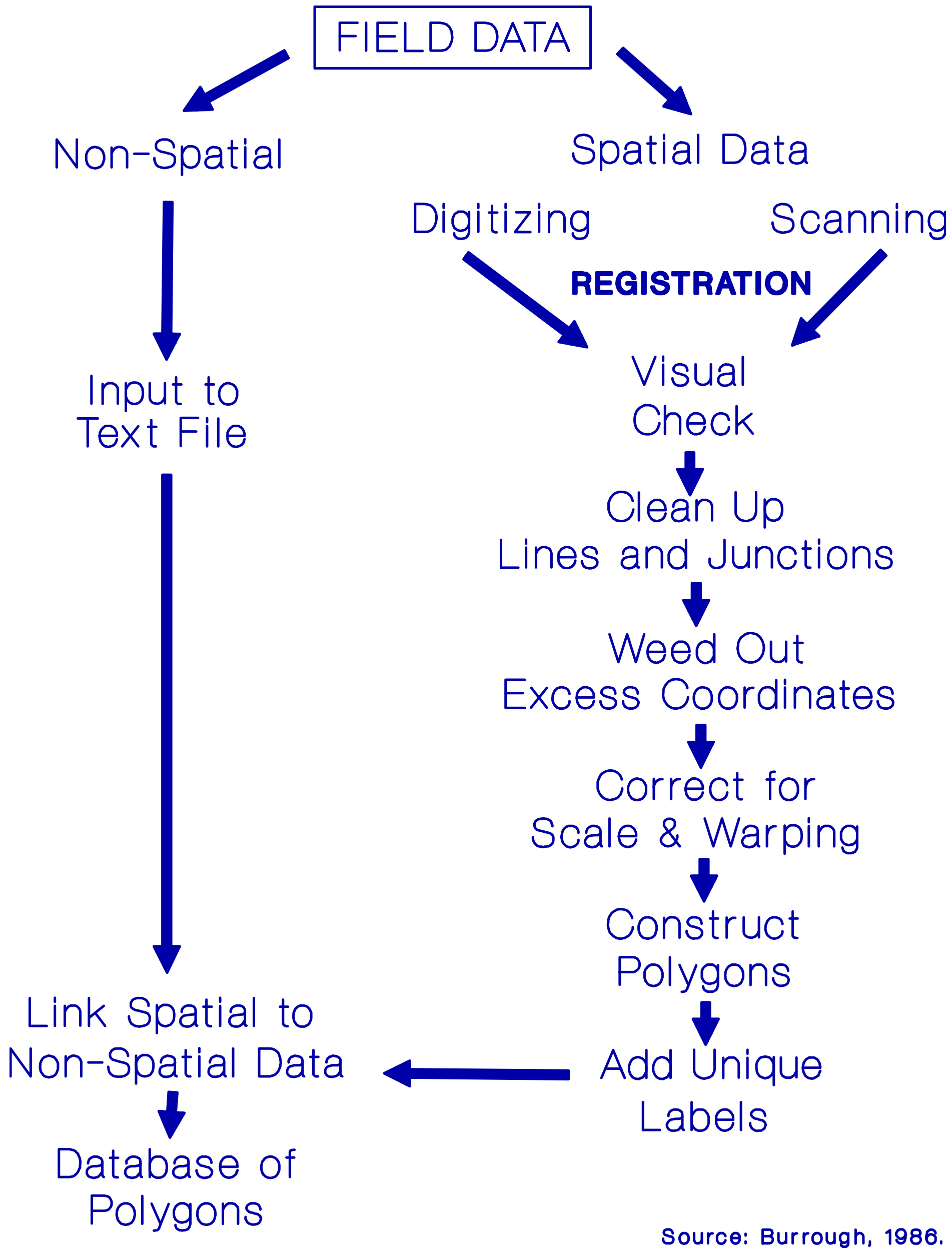
Six clear steps stand out in the data editing and verification process for spatial data.[28] These are :

- **Visual review**. This is usually by check plotting.
- **Cleanup of lines and junctions**. This process is usually done by software first and interactive editing second.
- **Weeding of excess coordinates**. This process involves the removal of redundant vertices by the software for linear and/or polygonal features.
- **Correction for distortion and warping**. Most GIS software has functions for scale correction and *rubber sheeting*. However, the distinct rubber sheet algorithm used will vary depending on the spatial data model, vector or raster, employed by the GIS. Some raster techniques may be more intensive than vector based algorithms.
- **Construction of polygons**. Since the majority of data used in GIS is polygonal, the construction of polygon features from lines/arcs is necessary. Usually this is done in conjunction with the topological building process.
- **The addition of unique identifiers or labels**. Often this process is manual. However, some systems do provide the capability to automatically build labels for a data layer.

[27] Burrough, 1986.

[28] Burrough, 1986.

These data verification steps occur after the data input stage and prior to or during the linkage of the spatial data to the attributes. Data verification ensures the integrity between the spatial and attribute data. Verification should include some brief querying of attributes and cross checking against known values. The adjoining figure presents the steps typically involved in creating a topologically correct set of vector polygons.



The typical steps involved in creating a topologically correct vector polygon database.

4.0 Data Storage and Retrieval

This chapter reviews the approaches for organizing and maintaining data in a GIS. The focus is on reviewing different techniques for storing spatial data. A brief review of data querying approaches for attribute data is also provided.

The second necessary component for a GIS is the data storage and retrieval subsystem. This subsystem organizes the data, both spatial and attribute, in a form which permits it to be quickly retrieved for updating, querying, and analysis. Most GIS software utilizes proprietary software for their spatial editing and retrieval system, and a database management system (DBMS) for their attribute storage. Typically, an internal data model is used to store primary attribute data associated with the topological definition of the spatial data. Most often these internal database tables contain primary columns such as area, perimeter, length, and internal feature id number. Often thematic attribute data is maintained in an external DBMS that is linked to the spatial data via the internal database table.

4.1 Organizing Data for Analysis

Most GIS software organizes spatial data in a thematic approach that categorizes data in vertical *layers*. The definition of layers is fully dependent on the organization's requirements. Typical layers used in natural resource management agencies or companies include forest cover, soil classification, elevation, road network (access), ecological areas, hydrology, etc.

Spatial data layers are commonly input one at a time, e.g. forest cover. Accordingly, attribute data is entered one layer at a time. Depending on the attribute data model used by the data storage subsystem data must be organized in a format that will facilitate the manipulation and analysis tasks that will be required. Most often, the spatial and attribute data may be entered at different times and linked together later. However, this is fully dependent on the source of data.

The clear identification of the requirements for any GIS project is necessary before any data input procedures, and/or layer definitions, should occur.

It is mandatory that GIS users fully understand their needs before undertaking a GIS project.

Experience has shown that a less than complete understanding of the needs and processing tasks required for a specific project, greatly increases the time required to complete the project, and ultimately affects the quality and reliability of the derived GIS product(s).

4.1.1 Spatial Data Layers - Vertical Data Organization

In most GIS software data is organized in themes as *data layers*. This approach allows data to be input as separate themes and overlaid based on analysis requirements. This can conceptualized as vertical layering the characteristics of the earth's surface. The overlay concept is so natural to cartographers and natural resource specialists that it has been built into the design of most CAD vector systems as well.[29] The overlay/layer approach used in CAD systems is used to separate major classes of spatial features. This concept is also used to logically order data in most GIS software. The terminology may differ between GIS software, but the approach is the same. A variety of terms are used to define data layers in commercial GIS software. These include themes, coverages, layers, levels, objects, and feature classes. Data layer and theme are the most common and the least proprietary to any particular GIS software and accordingly, as used throughout the book.

In any GIS project a variety of data layers will be required. These must be identified before the project is started and a priority given to the input or digitizing of the spatial data layers. This is mandatory, as often one data layer contains features that are coincident with another, e.g. lakes can be used to define polygons within the forest inventory data layer. Data layers are commonly defined based on the needs of the user and the availability of data. They are completely user definable.

When considering the physical requirements of the GIS software it is important to understand that two types of data are required for each layer., attribute and spatial data. Commonly, data layers are input into the GIS one layer at a time. As well, often a data layer is completely loaded, e.g. graphic conversion, editing, topological building, attribute conversion, linking, and verification, before the next data layer is started. Because there are several steps involved in completely loading a data layer it can become very confusing if many layers are loaded at once.

The proper identification of layers prior to starting data input is critical. The identification of data layers is often achieved through a *user needs analysis*. The user needs analysis performs several functions including :

- identifying the users;
- educating users with respect to GIS needs;
- identifying information products;
- identifying data requirements for information products;
- priorizing data requirements and products; and
- determining GIS functional requirements.

Often a user needs assessment will include a review of existing operations, e.g. sometimes called a *situational assessment*, and a *cost-benefit analysis*. The cost-benefit process is well established in conventional data processing and serves as the mechanism to justify acquisition of hardware and software. It defines and compares costs against potential benefits. Most institutions will require this step before a GIS acquisition can be undertaken.

For illustration purposes the following list of sample data layers is provided. It represents

[29] Burrough, 1986.

a sample set of data layers that might be used for a typical operational forestry GIS project. This listing is not intended to be comprehensive, but merely provides a simply example of different layers.

Thematic Data

- Forest inventory stands;
- Soil survey;
- Wildlife habitat;
- Physical land classification;
- Ecological land classification; and
- Elevation (and related derivatives).

Positional Reference (Base Map) Data

- Survey / legal description fabric;
- Hydrography data (lakes and rivers);
- Transportation network (access) data; and
- Administrative boundaries (e.g. forest management units, tenure)

Most GIS projects integrate data layers to create derived themes or layers that represent the result of some calculation or geographic model, e.g. forest merchantability, land use suitability, etc. Derived data layers are completely dependant on the aim of the project.

Each data layer would be input individually and topologically integrated to create combined data layers. Based on the data model, e.g. vector or raster, and the topological structure, selected data analysis functions could be undertaken. It is important to note that in vector based GIS software the topological structure defined can only be traversed by means of unique labels to every feature.

4.1.2 Spatial Indexing - Horizontal Data Organization

The proprietary organization of data layers in a horizontal fashion within a GIS is known as *spatial indexing*. Spatial indexing is the method utilized by the software to store and retrieve spatial data. A variety of different strategies exist for speeding up the spatial feature retrieval process within a GIS software product. Most involve the partitioning of the geographic area into manageable subsets or *tiles*. These tiles are then indexed mathematically, e.g. by quadtrees, by R (rectangle) trees, to allow for quick searching and retrieval when querying is initiated by a user. Spatial indexing is analygous to the definition of map sheets, except that specific indexing techniques are used to access data across map sheet (tile) boundaries. This is done simply to improve query performance for large data sets that span multiple map sheets, and to ensure data integrity across map sheet boundaries.

The method and process of spatial indexing is usually transparent to the user. However, it becomes very important especially when large data sets are utilized. The notion of spatial indexing has become increasingly important in the design of GIS software over the last few years, as larger scale applications have been initiated using GIS technology. Users have found that often the response time in querying very large data sets is unacceptably slow. GIS software vendors have responded by developing sophisticated algorithms to index and retrieve spatial data. It is important to note that raster systems, by the nature of their data structure, do not typically require a spatial indexing method. The raster approach imposes regular, readily addressable partitions on the data universe intrinsically with its data structure. Accordingly, spatial indexing is usually not required. However, the more sophisticated vector GIS does require a method to quickly retrieve spatial objects.

The horizontal indexing of spatial data within GIS software involves several issues. These concern the extent of the spatial indexing approach. They include :

- the use of a librarian subsystem to organize data for users;
- the requirement for a formal definition of layers;
- the need for feature coding within themes or layers; and
- requirements to maintain data integrity through transaction control, e.g. the locking of selected spatial tiles (or features) when editing is being undertaken by a permitted user.

While all these issues need not be satisfied for spatial indexing to occur, they are important aspects users should consider when evaluating GIS software.

While the spatial indexing method is usually not the selling point of any GIS, users should consider these requirements, especially if very large data sets, e.g. 10,000 + polygons, are to be the norm in their applications, and a vector data model is to be employed.

4.2 Editing and Updating of Data

Perhaps the primary function in the data storage and retrieval subsystem involves the editing and updating of data. Frequently, the following data editing capabilities are required :

- interactive editing of spatial data;
- interactive editing of attribute data;
- the ability to add, manipulate, modify, and delete both spatial features and attributes (independently or simultaneously) ; and the
- ability to edit selected features in a batch processing mode.

Updating involves more than the simple editing of features. Updating implies the resurvey and processing of new information.[30] The updating function is of great importance during any GIS project. The life span of most digital data can range anywhere from 1 to 10 years. Commonly, digital data is valid for 5 to 10 years. The lengthy time span is due to the intensive task of data capture and input. However, often periodic data updates are required. These

[30] Burrough, 1986.

frequently involve an increased accuracy and/or detail of the data layer. Changes in classification standards and procedures may necessitate such updates. Updates to a forest cover data layer to reflect changes from a forest fire burn or a harvest cut are typical examples.

Many times data updates are required based on the results of a derived GIS product. The generation of a derived product may identify blatant errors or inappropriate classes for a particular layer. When this occurs updating of the data is required. In this situation the GIS operator usually has some previous experience or knowledge of the study area.

Commonly **the data update process is a result of a physical change in the geographic landscape**. Forest fires are a prime example. With this type of update new features are usually required for the data layer, e.g. burn polygons. As well, existing features are altered, e.g. forest stands that were affected. There is a strong requirement for a **historical record keeping** capability with this type of update process. Users should be aware of this requirement and design their database organization to accommodate such needs. Depending on the particular GIS, the update process may involve some data manipulation and analysis functions.

4.3 Data Retrieval and Querying

The ability to query and retrieve data based on some user defined criteria is a necessary feature of the data storage and retrieval subsystem.

Data retrieval involves the capability to easily select data for graphic or attribute editing, updating, querying, analysis and/or display.

The ability to retrieve data is based on the unique structure of the DBMS and command interfaces are commonly provided with the software. Most GIS software also provides a programming subroutine library, or macro language, so the user can write their own specific data retrieval routines if required.

Querying is the capability to retrieve data, usually a data subset, based on some user defined formula. These data subsets are often referred to as *logical views*. Often the querying is closely linked to the data manipulation and analysis subsystem. Many GIS software offerings have attempted to standardize their querying capability by use of a *Standard Query Language (SQL)*. This is especially true with systems that make use of an external relational DBMS. Through the use of SQL, GIS software can interface to a variety of different DBMS packages. This approach provides the user with the flexibility to select their own DBMS. This has direct implications if the organization has an existing DBMS that is being used for to satisfy other business requirements. Often it is desirable for the same DBMS to be utilized in the GIS applications. This notion of integrating the GIS software to utilize an existing DBMS through standards is referred to as *corporate or enterprise GIS*. With the migration of GIS technology from being a research tool to being a decision support tool there is a requirement for it to be totally integrated with existing corporate activities, including accounting, reporting, and business functions.

There is a definite trend in the GIS marketplace towards a generic interface with external relational DBMS's. The use of an external DBMS, linked via a SQL interface, is becoming the norm. A flexibility as such is a strong selling point for any GIS. SQL is quickly becoming a standard in the GIS software marketplace.

5.0 Data Manipulation and Analysis

This chapter reviews data manipulation and analysis capabilities within a GIS. The focus is on reviewing spatial data analytical functions. This chapter categorizes analytical functions within a GIS and will be of most interest to technical staff and GIS operators.

The major difference between GIS software and CAD mapping software is the provision of capabilities for transforming the original spatial data in order to be able to answer particular queries. Some transformation capabilities are common to both GIS and CAD systems, however, GIS software provides a larger range of analysis capabilities that will be able to operate on the topology or spatial aspects of the geographic data, on the non-spatial attributes of these data, or on both.[31]

The main criteria used to define a GIS is its capability to transform and integrate spatial data.

5.1 Manipulation and Transformation of Spatial Data

The maintenance and transformation of spatial data concerns the ability to input, manipulate, and transform data once it has been created. While many different interpretations exist with respect to what constitutes these capabilities some specific functions can be identified.[32] These are reviewed below.

5.1.1 Coordinate Thinning

Coordinate thinning involves the *weeding* or reduction of coordinate pairs, e.g. X and Y, from arcs. This function is often required when data has been captured with too many vertices for the linear features. This can result in redundant data and large data volumes. The weeding of coordinates is required to reduce this redundancy.

The thinning of coordinates is also required in the map generalization process of *linear simplification*. Linear simplification is one component of generalization that is required when data from one scale, e.g. 1:20,000, is to be used and integrated with data from another scale, e.g. 1:100,000. Coordinate thinning is often done on features such as contours, hydrography, and forest stand boundaries.

[31] Burrough, 1986.

[32] Aronoff presents an excellent categorization of different manipulation and analysis functions.

5.1.2 Geometric Transformations

This function is concerned with the registering of a data layer to a common coordinate scheme. This usually involves registering selected data layers to a standard data layer already registered. The term *rubber sheeting* is often used to describe this function. Rubber sheeting involves stretching one data layer to meet another based on predefined control points of known locations. Two other functions may be categorized under geometric transformations. These involve *warping* a data layer stored in one data model, either raster or vector, to another data layer stored in the opposite data model. For example, often classified satellite imagery may require warping to fit an existing forest inventory layer, or a poor quality vector layer may require warping to match a more accurate raster layer.

5.1.3 Map Projection Transformations

This functionality concerns the transformation of data in geographic coordinates for an existing map projection to another map projection. Most GIS software requires that data layers must be in the same map projection for analysis. Accordingly, if data is acquired in a different projection than the other data layers it must be transformed. Typically 20 or more different map projections are supported in a GIS software offering.

5.1.4 Conflation - Sliver Removal

Conflation is formally defined as the procedure of reconciling the positions of corresponding features in different data layers.[33] More commonly this is referred to as *sliver removal*. Often two layers that contain the same feature, e.g. soils and forest stands both with a specific lake, do not have exactly the same boundaries for that feature, e.g. the lake. This may be caused by a lack of coordination or data prioritization during digitizing or by a number of different manipulation and analysis techniques. When the two layers are combined, e.g. normally in polygon overlay, they will not match precisely and small sliver polygons will be created. Conflation is concerned with the process for removing these slivers and reconciling the common boundary.

There are several approaches for sliver removal. Perhaps the most common is allowing the user to define a priority for data layers in combination with a tolerance value. Considering the soils and forest stand example the user could define a layer that takes precedence, e.g. forest stands, and a size tolerance for slivers. After polygon overlay if a polygon is below the size tolerance it is classified a sliver. To reconcile the situation the arcs of the data layer that has higher priority will be retained and the arcs of the other data layer will be deleted. Another approach is to simply divide the sliver down the centre and collapse the arcs making up the boundary. The important point is that all GIS software must have the capability to resolve slivers. Remember that it is generally much less expensive to reconcile maps manually in the map preparation and digitizing stage than afterwards.

[33] Aronoff, 1989.

5.1.5 Edge Matching

Edge matching is simply the procedure to adjust the position of features that extend across typical map sheet boundaries. Theoretically data from adjacent map sheets should meet precisely at map edges. However, in practice this rarely occurs. Misalignment of features can be caused by several factors including digitizing error, paper shrinkage of source maps, and errors in the original mapping. Edge matching always requires some interactive editing. Accordingly, GIS software differs considerably in the degree of automation provided.

5.1.6 Interactive Graphic Editing

Interactive graphic editing functions involve the addition, deletion, moving, and changing of the geographic position of features. Editing should be possible at any time. Most graphic editing occurs during the data compilation phase of any project. Remember typically 60 to 70 % of the time required to complete any project involves data compilation. Accordingly, the level of sophistication and ease of use of this capability is vitally important and should be rated highly by those evaluating GIS software. Many of the editing that is undertaken involves the cleaning up of topological errors identified earlier. The capability to *snap* to existing elements, e.g. nodes and arcs, is critical.

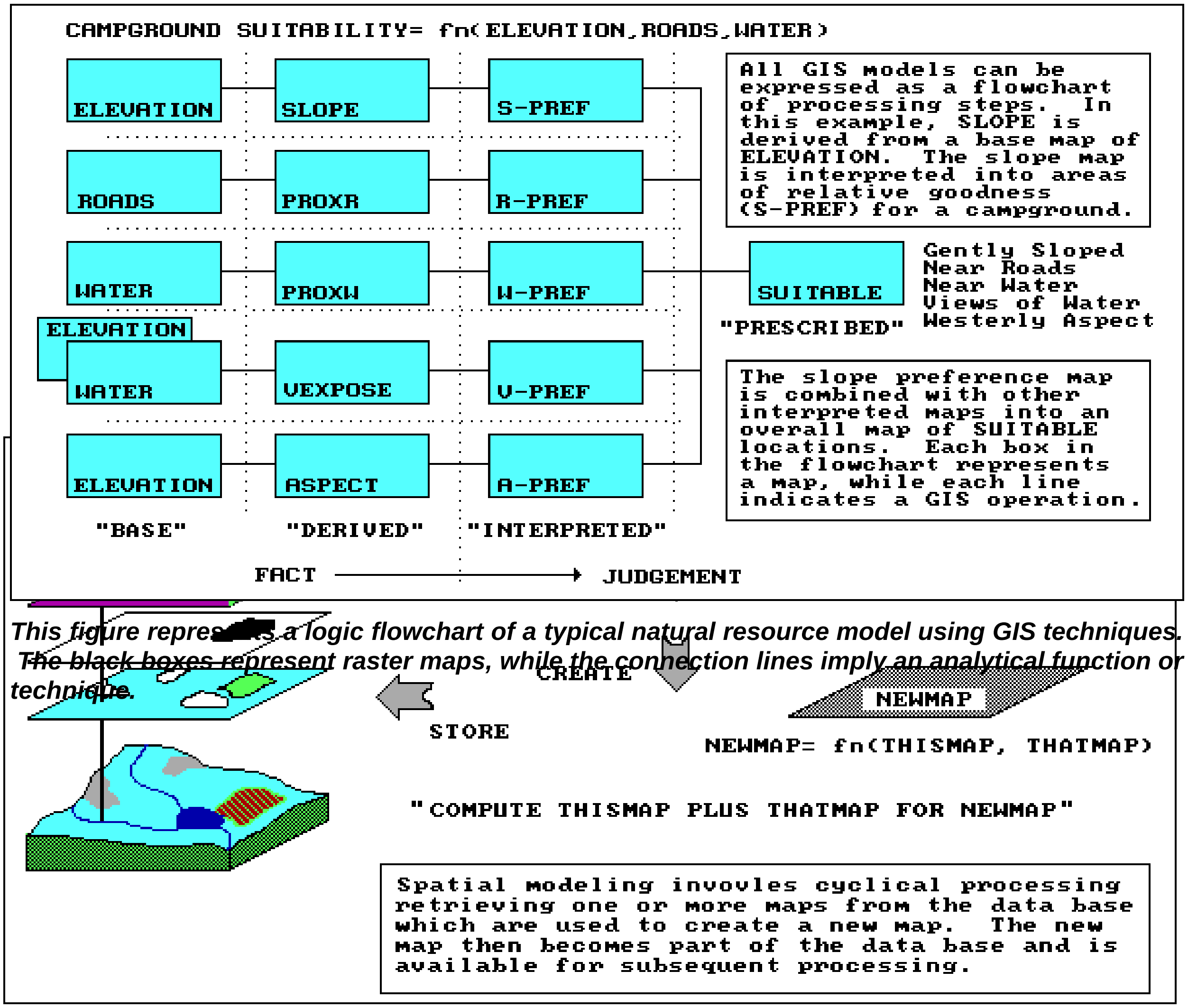
The functionality of graphic editing does not differ greatly across GIS software offerings. However, the user interface and ease of use of the editing functions usually does. Editing within a GIS software package should be as easy as using a CAD system. A cumbersome or incomplete graphic editing capability will lead to much frustration by the users of the software.

5.2 Integration and Modelling of Spatial Data

The integration of data provides the ability to ask complex spatial questions that could not be answered otherwise. Often, these are inventory or locational questions such as *how much ?* or *where ?*. Answers to locational and quantitative questions require the combination of several different data layers to be able to provide a more complete and realistic answer. The ability to combine and integrate data is the backbone of GIS.

Often, applications do require a more sophisticated approach to answer complex spatial queries and *what if ?* scenarios. The technique used to solve these questions is called *spatial modelling*. Spatial modelling infers the use of spatial characteristics and methods in manipulating data. Methods exist to create an almost unlimited range of capabilities for data analysis by stringing together sets of primitive analysis functions. While some explicit analytical models do exist, especially in natural resource applications, most *modelling formulae* (models) are determined based on the needs of a particular project. The capability to undertake complex modelling of spatial data, on an ad hoc basis, has helped to further the resource specialists understanding of the natural environment, and the relationship between selected characteristics of that environment.

The use of GIS spatial modelling tools in several traditional resource activities has helped to quantify processes and define models for deriving analysis products. This is particularly true in the area of resource planning and inventory compilation. Most GIS users are able to better organize their applications because of their interaction with, and use of, GIS technology. The utilization of spatial modelling techniques requires a comprehensive



This figure represents a logic flowchart of a typical natural resource model using GIS techniques. The black boxes represent raster maps, while the connection lines imply an analytical function or technique.

understanding of the data sets involved, and the analysis requirements.

The critical function for any GIS is the integration of data.[34]

The raster data model has become the primary spatial data source for analytical modeling with GIS. The raster data model is well suited to the quantitative analysis of numerous data layers. To facilitate these raster modeling techniques most GIS software employs a separate module specifically for cell processing.

[34] Anderson, Starr. 1984.

5.3 Integrated Analytical Functions in a GIS

Most GIS's provide the capability to build complex models by combining primitive analytical functions. Systems vary as to the complexity provided for spatial modelling, and the specific functions that are available. However, most systems provide a standard set of primitive analytical functions that are accessible to the user in some logical manner. Aronoff identifies four categories of GIS analysis functions. These are :

- Retrieval, Reclassification, and Generalization;
- Overlay Techniques;
- Neighbourhood Operations; and
- Connectivity Functions.

The range of analysis techniques in these categories is very large. Accordingly, this section of the book focuses on providing an overview of the fundamental primitive functions that are most often utilized in spatial analyses.[35]

5.3.1 Retrieval, Reclassification and Generalization

Perhaps the initial GIS analysis that any user undertakes is the *retrieval* and/or *reclassification* of data. Retrieval operations occur on both spatial and attribute data. Often data is selected by an attribute subset and viewed graphically. Retrieval involves the selective search, manipulation, and output of data without the requirement to modify the geographic location of the features involved.

Reclassification involves the selection and presentation of a selected layer of data based on the classes or values of a specific attribute, e.g. cover group. It involves looking at an attribute, or a series of attributes, for a single data layer and classifying the data layer based on the range of values of the attribute. Accordingly, features adjacent to one another that have a common value, e.g. cover group, but differ in other characteristics, e.g. tree height, species, will be treated and appear as one class. In raster based GIS software, numerical values are often used to indicate classes. Reclassification is an attribute generalization technique. Typically this function makes use of polygon patterning techniques such as crosshatching and/or color shading for graphic representation.

In a vector based GIS, boundaries between polygons of common reclassed values should be *dissolved* to create a cleaner map of homogeneous continuity. Raster reclassification intrinsically involves boundary dissolving. The dissolving of map boundaries based on a specific attribute value often results in a new data layer being created. This is often done for visual clarity in the creation of derived maps. Almost all GIS software provides the capability to easily dissolve boundaries based on the results of a reclassification. Some systems allow the user to

[35] A more detailed review of these categories can be found in Aronoff, 1989.

create a new data layer for the reclassification while others simply dissolve the boundaries during data output.

One can see how the querying capability of the DBMS is a necessity in the reclassification process. The ability and process for displaying the results of reclassification, a map or report, will vary depending on the GIS. In some systems the querying process is independent from data display functions, while in others they are integrated and querying is done in a graphics mode. The exact process for undertaking a reclassification varies greatly from GIS to GIS. Some will store results of the query in *query sets* independent from the DBMS, while others store the results in a newly created attribute column in the DBMS. The approach varies drastically depending on the architecture of the GIS software.

5.3.2 Topological Overlay

The capability to overlay multiple data layers in a vertical fashion is the most required and common technique in geographic data processing. In fact, the use of a topological data structure can be traced back to the need for overlaying vector data layers. With the advent of the concepts of mathematical topology *polygon overlay* has become the most popular geoprocessing tool, and the basis of any functional GIS software package.

Topological overlay is predominantly concerned with overlaying polygon data with polygon data, e.g. soils and forest cover. However, there are requirements for overlaying point, linear, and polygon data in selected combinations, e.g. point in polygon, line in polygon, and polygon on polygon are the most common. Vector and raster based software differ considerably in their approach to topological overlay.

Raster based software is oriented towards arithmetic overlay operations, e.g. the addition, subtraction, division, multiplication of data layers. The nature of the *one attribute map* approach, typical of the raster data model, usually provides a more flexible and efficient overlay capability. The raster data model affords a strong numerically modelling (quantitative analysis) modelling capability. Most sophisticated spatial modelling is undertaken within the raster domain.

In vector based systems topological overlay is achieved by the creation of a new topological network from two or more existing networks. This requires the rebuilding of topological tables, e.g. arc, node, polygon, and therefore can be time consuming and CPU intensive. The result of a topological overlay in the vector domain is a new topological network that will contain attributes of the original input data layers. In this way selected queries can then be undertaken of the original layer, e.g. soils and forest cover, to determine where specific situations occur, e.g. deciduous forest cover where drainage is poor.

Most GIS software makes use of a consistent logic for the overlay of multiple data layers. The rules of *Boolean logic* are used to operate on the attributes and spatial properties of geographic features. Boolean algebra uses the operators AND, OR, XOR, NOT to see whether a particular condition is true or false. Boolean logic represents all possible combinations of spatial interaction between different features. The implementation of Boolean operators is often transparent to the user.

To date the primary analysis technique used in GIS applications, vector and raster, is the topological overlay of selected data layers.

Generally, GIS software implements the overlay of different vector data layers by combining the spatial and attribute data files of the layers to create a new data layer. Again, different GIS software utilize varying approaches for the display and reporting of overlay results. Some systems require that topological overlay occur on only two data layers at a time, creating a third layer. This *pairwise* approach requires the nesting of multiple overlays to generate a final overlay product, if more than two data layers are involved. This can result in numerous intermediate or temporary data layers. Some systems create a complete topological structure at the data verification stage, and the user merely submits a query string for the combined topological data. Other systems allow the user to overlay multiple data layers at one time. Each approach has its drawbacks depending on the application and the nature of the implementation. Determining the most appropriate method is based on the type of application, practical considerations such as data volumes and CPU power, and other considerations such personnel and time requirements. Overall, the flexibility provided to the operator and the level of performance varies widely among GIS software offerings.

5.3.3 Neighbourhood Operations

Neighbourhood operations evaluate the characteristics of an area surrounding a specific location. Virtually all GIS software provides some form of neighbourhood analysis. A range of different neighbourhood functions exist. The analysis of topographic features, e.g. the relief of the landscape, is normally categorized as being a neighbourhood operation. This involves a variety of *point interpolation* techniques including slope and aspect calculations, contour generation, and *Thiessen polygons*. Interpolation is defined as the method of predicting unknown values using known values of neighbouring locations.[36] Interpolation is utilized most often with point based elevation data.

Elevation data usually takes the form of irregular or regular spaced points. Irregularly space points are stored in a Triangular Irregular Network (*TIN*). A TIN is a vector topological network of triangular facets generated by joining the irregular points with straight line segments. The TIN structure is utilized when irregular data is available, predominantly in vector based systems. TIN is a vector data model for 3-D data.

An alternative in storing elevation data is the regular point Digital Elevation Model (*DEM*). The term DEM usually refers to a grid of regularly space elevation points. These points are usually stored with a raster data model. Most GIS software offerings provide three dimensional

[36] Aronoff, 1989.

analysis capabilities in a separate module of the software. Again, they vary considerably with respect to their functionality and the level of integration between the 3-D module and the other more typical analysis functions.

Without doubt the most common neighbourhood function is *buffering*. Buffering involves the ability to create distance buffers around selected features, be it points, lines, or areas. Buffers are created as polygons because they represent an area around a feature. Buffering is also referred to as *corridor* or *zone generation* with the raster data model. Usually, the results of a buffering process are utilized in a topological overlay with another data layer. For example, to determine the volume of timber within a selected distance of a cutline, the user would first buffer the cutline data layer. They would then overlay the resultant buffer data layer, a buffer polygon, with the forest cover data layer in a clipping fashion. This would result in a new data layer that only contained the forest cover within the buffer zone. Since all attributes are maintained in the topological overlay and buffering processes, a map or report could then be generated.

Buffering is typically used with point or linear features. The generation of buffers for selected features is frequently based on a distance from that feature, or on a specific attribute of that feature. For example, some features may have a greater zone of influence due to specific characteristics, e.g. a primary highway would generally have a greater influence than a gravel road. Accordingly, different size buffers can be generated for features within a data layer based on selected attribute values or feature types.

5.3.4 Connectivity Analysis

The distinguishing feature of connectivity operations is that they use functions that accumulate values over an area being traversed.[37] Most often these include the analysis of surfaces and networks. Connectivity functions include *proximity analysis*, *network analysis*, spread functions, and three dimensional surface analysis such as *visibility* and *perspective viewing*. This category of analysis techniques is the least developed in commercial GIS software. Consequently, there is often a great difference in the functionality offered between GIS software offerings. Raster based systems often provide the more sophisticated surface analysis capabilities while vector based systems tend to focus on linear network analysis capabilities. However, this appears to be changing as GIS software becomes more sophisticated, and multi-disciplinary applications require a more comprehensive and integrated functionality. Some GIS offerings provide both vector and raster analysis capabilities. Only in these systems will one fund a full range of connectivity analysis techniques.

Proximity analysis techniques are primarily concerned with the proximity of one feature to another. Usually *proximity* is defined as the ability to identify any feature that is near any other feature based on location, attribute value, or a specific distance. A simple example is identifying all the forest stands that are within 100 metres of a gravel road, but not necessarily adjacent to it. It is important to note that neighbourhood buffering is often categorized as being a proximity analysis capability. Depending on the particular GIS software package, the data model

[37] Aronoff, 1989.

employed, and the operational architecture of the software it may be difficult to distinguish proximity analysis and buffering.

The identification of *adjacency* is another proximity analysis function. Adjacency is defined as the ability to identify any feature having certain attributes that exhibit adjacency with other selected features having certain attributes. A typical example is the ability to identify all forest stands of a specific type, e.g. specie, adjacent to a gravel road.

Network analysis is a widely used analysis technique. Network analysis techniques can be characterized by their use of *feature networks*. Feature networks are almost entirely comprised of linear features. Hydrographic hierarchies and transportation networks are prime examples. Two example network analysis techniques are the *allocation of values* to selected features within the network to determine capacity zones, and the determination of *shortest path* between connected points or nodes within the network based on attribute values. This is often referred to as *route optimization*. Attribute values may be as simple as minimal distance, or more complex involving a model using several attributes defining rate of flow, impedance, and cost.

Three dimensional analysis involves a range of different capabilities. The most utilized is the generation of perspective surfaces. Perspective surfaces are usually represented by a wire frame diagram reflecting profiles of the landscape, e.g. every 100 metres. These profiles viewed together, with the removal of hidden lines, provide a three dimensional view. As previously identified, most GIS software packages offer 3-D capabilities in a separate module. Several other functions are normally available.

These include the following functions :

- user definable *vertical exaggeration*, *viewing azimuth*, and *elevation angle*;
- identification of *viewsheds*, e.g. seen versus unseen areas;
- the *draping* of features, e.g. point, lines, and shaded polygons onto the perspective surface;
- generation of shaded relief models simulating illumination;
- generation of cross section profiles;
- presentation of symbology on the 3-D surface; and
- line of sight perspective views from user defined viewpoints.

While the primitive analytical functions have been presented the reader should be aware that a wide range of more specific and detailed capabilities do exist.

The overriding theme of all GIS software is that the analytical functions are totally integrated with the DBMS component. This integration provides the necessary foundation for all analysis techniques.

6.0 Sample Forestry Applications

This chapter presents some sample forestry applications. The focus is on illustrating the use of GIS analytical functions to address typical operational forestry requirements.

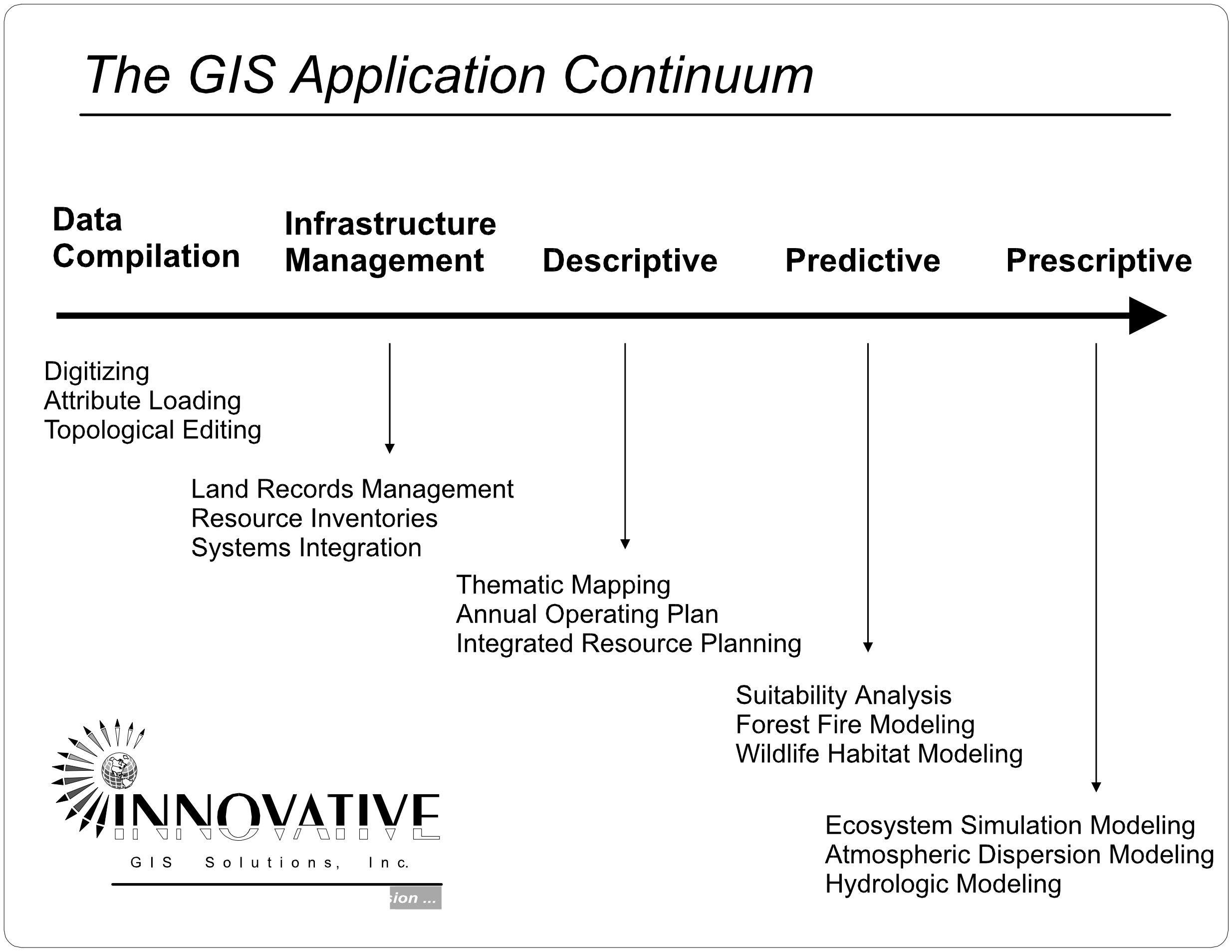
To illustrate some of the primitive analytical functions **three** sample applications are presented. These applications were developed in conjunction with foresters and resource specialists, and represent realistic scenarios within the forestry industry. However, many of the models utilized have been simplified as the focus is on illustrating the application of analytical GIS techniques. The sample applications are :

1. Resource Capabilities - Reclassifying Forest Cover and Soil Cover
2. Forest Inventory Updating - Topological Overlay to Update Forest Cover
3. DEM Modeling - Using Forest Cover with a DEM for Determining Merchantability

The first two examples illustrate reclassification and topological overlay. The study area comprises 12 survey sections crossing two legal townships in north eastern Alberta. Maps and reports are presented.

The third example represents a more integrated analytical approach. It makes use of procedures used in the first three examples but also includes the use of a Digital Elevation Model (DEM) data. The study area is also located in north western Alberta. This area has very steep terrain and is an excellent example for using elevation data. It focuses on the identification of stand merchantability based on selected attributes of the forest cover in combination with characteristics of the terrain. Maps and reports are presented.

It is important to note that the examples presented here reflect initial uses of GIS technology in the *GIS application continuum*. The GIS application continuum reflects the logical progression of GIS use in applying GIS technology to land management issues and problems. Initial uses address simple data compilation and infrastructure management requirements, e.g. such as forest inventory maintenance. As more is understood about the data and technology the user migrates into more descriptive uses, e.g. thematic mapping etc. Depending on specific operational requirements this can lead into predictive modeling, e.g. wildlife habitat modeling, etc., and lastly into prescriptive analysis, e.g. simulation modeling, wood supply analysis, etc. Most users attain descriptive use of the technology fairly quickly, however the hurdle to applying GIS in a predictive and prescriptive manner may be lengthy and complex depending on requirements. In most cases the predictive and prescriptive application of GIS technology requires integration of GIS with other technologies, often discrete modeling programs written in FORTRAN or C. The GIS application continuum is presented in following figure.



The application continuum reflects a logical migration in use of GIS from simple data compilation and infrastructure management to more sophisticated analysis that often requires integration with external models and programs.

6.1 Example 1. Resource Capabilities

This sample application uses both a physical land classification (**SOIL**) data layer and a forest inventory (**FOREST**) data layer.

The **first** set of maps illustrates the distribution of polygons for each data layer. The **SOIL** layer contains **71** polygons. The **FOREST** layer contains **187** polygons. A list of the attribute columns for each data layer is also included. Remember, that in the vector data model an attribute database table is *linked* to each data set. Accordingly, a database table is created for SOIL and for FOREST. Reclassification is based on the attribute table columns using conventional DBMS query techniques. A description of the columns (fields) in the SOIL and FOREST attribute tables is presented below.

Attribute List - SOIL Data Layer

This list describes the schema of the SOIL attribute table. The AREA, PERIMETER and SOIL-ID columns are calculated automatically by the GIS software during the topological building process, e.g. in constructing clean topological polygons. The SOIL-ID represents the unique identifier assigned to each polygon.

Field	*Column Name*	*Description*
1	AREA	Area of polygon
2	PERIMETER	Perimeter of polygon
3	SOIL-ID	Unique identifier (label)
4	PARMAT	Parent material
5	SURFEX	Surface expression
6	SLOPE	Slope class
7	SURFTEXT	Surface soil texture
8	UNDERTEXT	Underlying soil texture
9	SOIL	Soil classification
10	DRAIN	Drainage class
11	SERIES	Soil series
12	SUBREG	Physiographic subregion

Attribute List - FOREST Data Layer

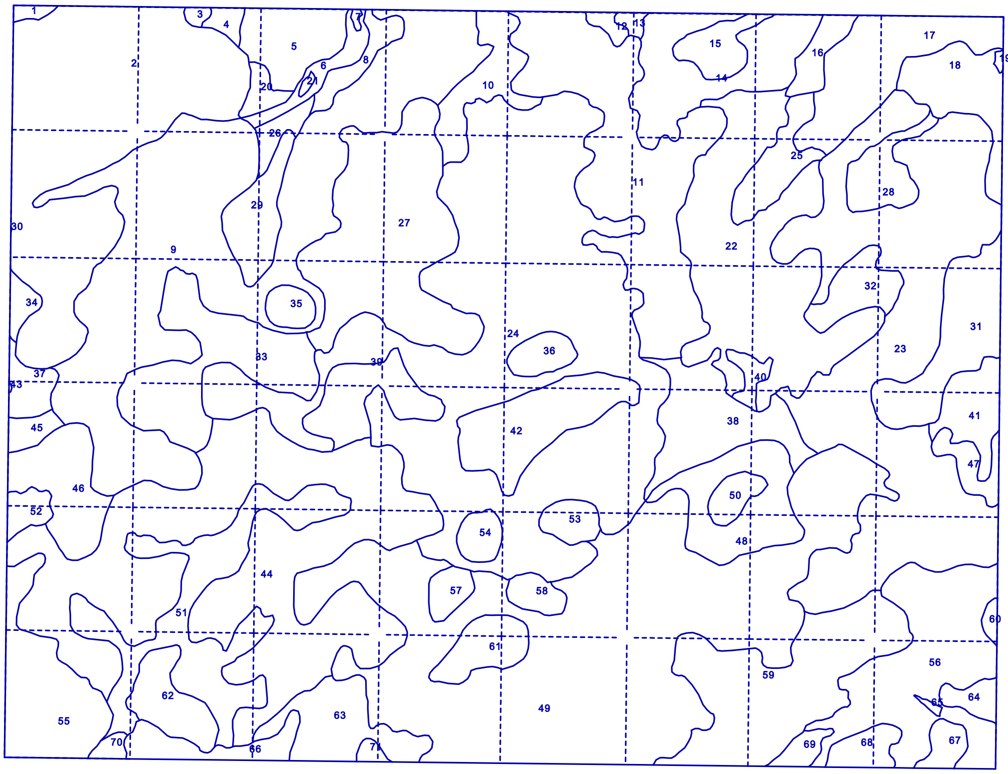
This list describes the schema of the FOREST attribute table. The AREA, PERIMETER and FOREST-ID columns are calculated automatically by the GIS software during the topological building process, e.g. in constructing clean topological polygons. The FOREST-ID represents the unique identifier assigned to each polygon.

Field	*Column Name*	*Description*
1	AREA	Area of stand
2	PERIMETER	Perimeter of stand
3	FOREST-ID	Unique identifier (label)
4	FORESTEST	Forest code
5	UNIT	Management unit
6	MANAGER	Manager
7	TWP	Township number
8	RGE	Range number
9	MER	Meridian
10	STANDNO	Stand number
11	STANDAL	Stand number alphanumeric code
12	STORY	Map story - under or over
13	DENSITY	Crown Density code
14	HEIGHT	Height code
15	SP1	Dominant species One code
16	SP2	Dominant species Two code

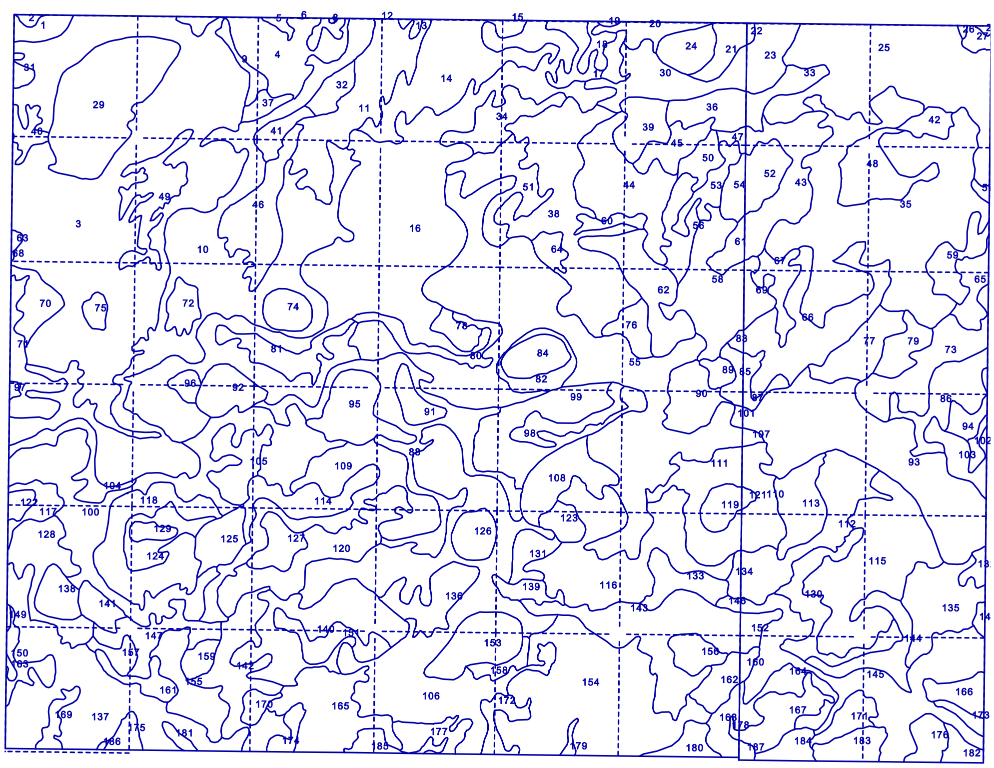
Attribute List cont'd - FOREST data layer

Field	*Column Name*	*Description*
17	SP3	Species Three code
18	SP4	Species Four code
19	COM	Commercialism code
20	VSR	Volume Sampling region (VSR)
21	CSP1	Coded species code one
22	CSP2	Coded species code two
23	CSP3	Coded species code three
24	CSP4	Coded species code four
25	CVGRP1	Cover group one
26	CVGRP2	Cover group two
27	SITE	Site index code
28	SLOPE	Slope code
29	PLANZONE	Planning zone
30	G/W	G/W zone code
31	OCEAN	Ocean drainage
32	BASIN	Drainage basin
33	WATERSHED	Watershed
34	SUBWATER	Sub watershed
35	ORIGIN	Origin of stand (year)
36	CON15	Live coniferous volume 15+/11 cm
37	DEC15	Live deciduous volume 15+/11
38	CON20	Live coniferous volume 19+/13
39	DEC20	Live deciduous volume 19+/13
40	CON25	Live coniferous volume 25+/15
41	DEC25	Live deciduous volume 15+/15
42	**ORIGMER**	**Origin class / merchantability**

Map 1 presents the SOIL polygons. Textual labels are identified for each polygon. The Alberta Township System (ATS) survey fabric quarter sections are also presented for reference as dashed lines. **Map 2** presents the FOREST polygons. Labels are identified for each polygon. ATS quarter sections are presented as dashed lines for reference.

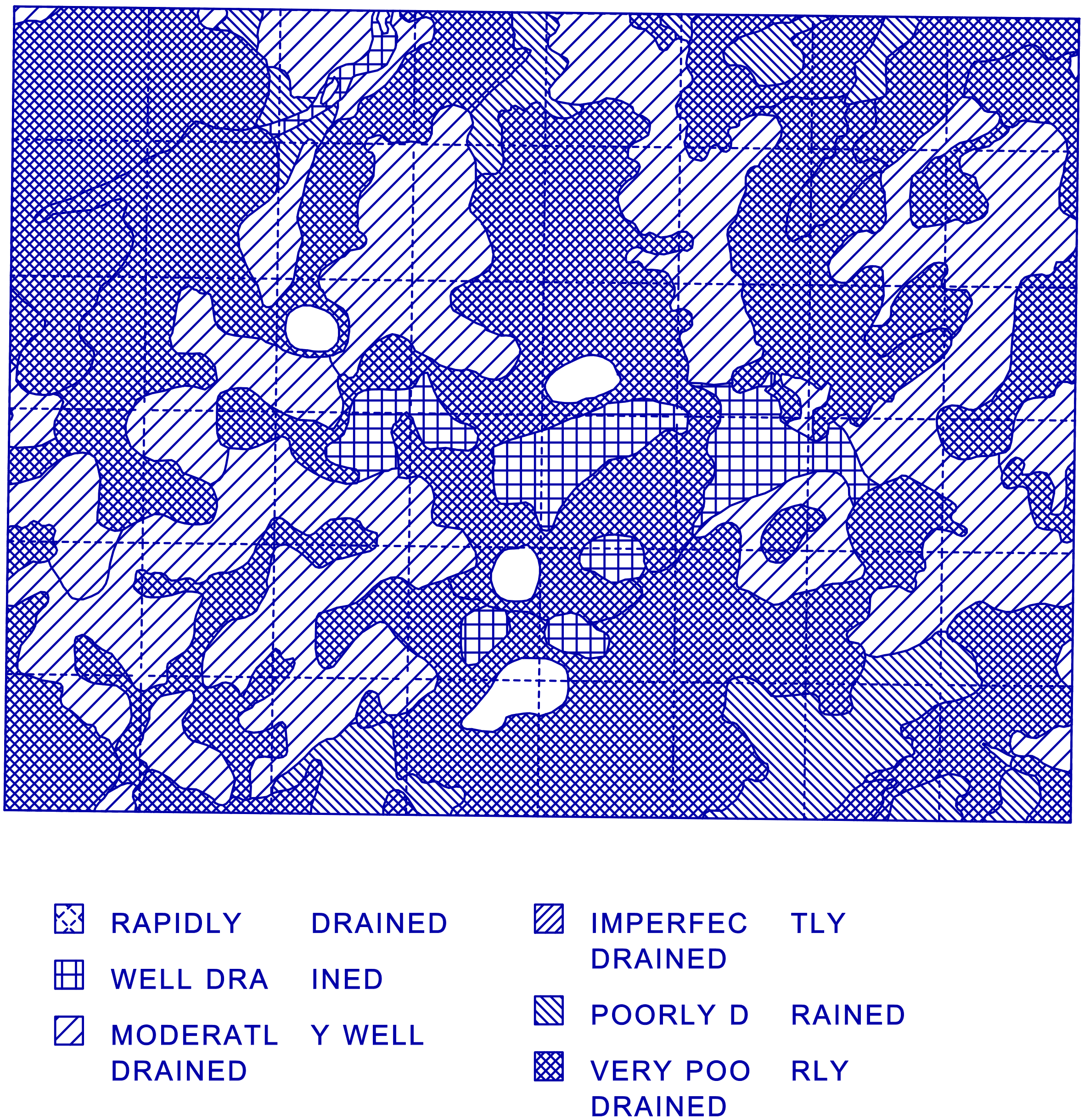


Example 1: Map 1 -SOIL polygons with labels. Unique numbers are used as labels to identify each soil polygon. A total of 71 polygons exists with this



Example 1: Map 2 - FOREST polygons with labels. Unique numbers are used as labels to identify each forest polygon. A total of 187 polygons exist

The second set of maps (maps 3-4) presents reclassifications of the SOIL data layer. **Map 3** is a simple reclassification based on original drainage class. Note that the feature polygons are the same, only the data content has changed. No *dissolving* of feature boundaries has occurred.



Example 1: Map 3 - A reclassification of the SOIL data layer by the DRAINAGE attribute. Each polygon has been color coded based on its drainage class attribute.

Map 4 represents a reclassification of selected attributes to derive a map of Soil Limitations. To derive this map two reclassification steps were undertaken. First, eight new attribute columns were added (appended) to the SOIL attribute table to permanently store the result of the first reclassification. A ninth attribute column was added to store the final limitation factor. These fields are defined below.

Field	*Column Name*	*Description*
13	ROADDRAIN	Drainage limitation for road location
14	ROADSLOPE	Slope limitation for road location
15	ROADSHRINK	Shrink-swell limitation for road location
16	ROADAASHO	AASHO index limitation for road location
17	ROADFROST	Frost heave limitation for road location
18	ROADROCK	Bedrock limitation for road location
19	ROADSTONE	Stoniness limitation for road location
20	ROADLIM	Final limitation rating for road location
21	ROADFACT	Final limitation factor for road location

Based on the individual characteristics of each polygon, the eight new limitation factors were derived for each polygon. Conventional database query techniques were used to distinguish unique features for the eight limitation classes. For example, to determine the ROADDRAIN attribute (drainage limitation for road location) for each polygon the following reclassification logic was employed.

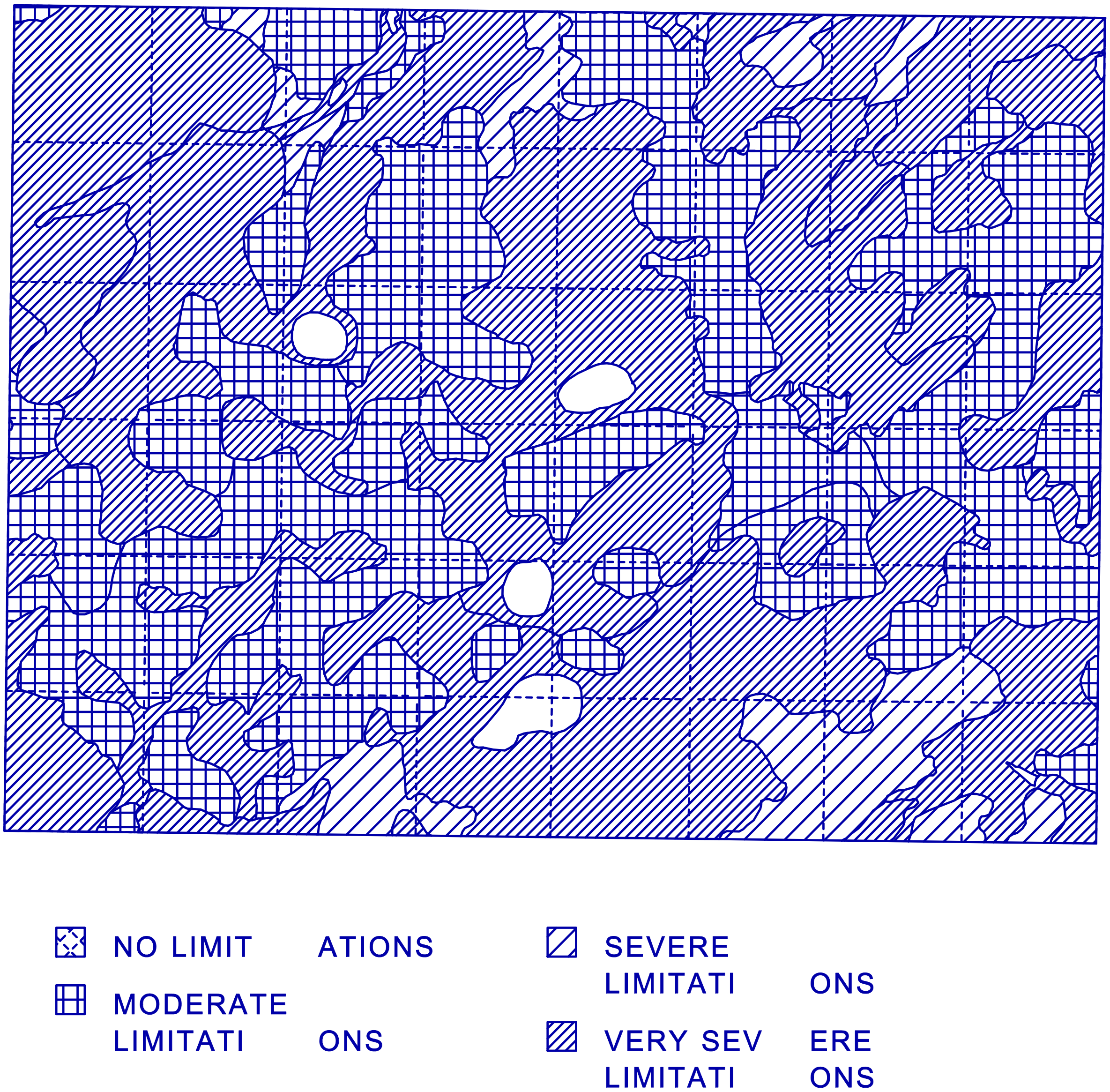
The existing drainage class attribute, DRAIN, was reclassified into groupings as follows:

if **DRAIN** equals **1 to 3** then **ROADDRAIN** equals '**N**'
if **DRAIN** equals **4** then **ROADDRAIN** equals '**M**'
if **DRAIN** equals **5** then **ROADDRAIN** equals '**S**'
if **DRAIN** equals **6** then **ROADDRAIN** equals '**VS**'

where **DRAIN** class **1** represents **RAPIDLY DRAINED**
DRAIN class **2** represents **WELL DRAINED**
DRAIN class **3** represents **MODERATELY WELL DRAINED**
DRAIN class **4** represents **IMPERFECTLY DRAINED**
DRAIN class **5** represents **POORLY DRAINED**
DRAIN class **6** represents **VERY POORLY DRAINED**

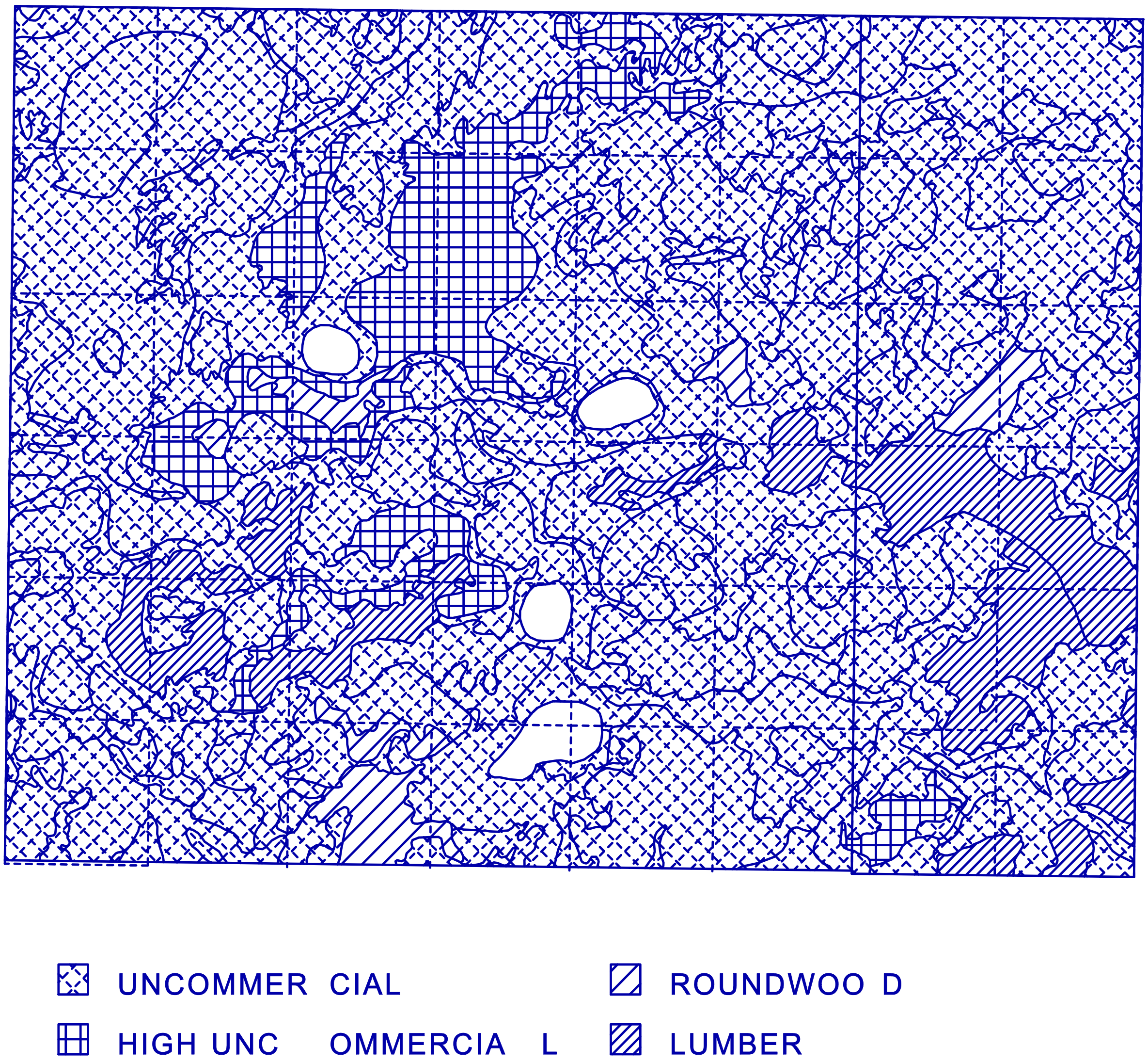
where '**N**' represents **NONE TO SLIGHT** limitations;
'**M**' represents **MODERATE** limitations;
'**S**' represents **SEVERE** limitations; and
'**VS**' represents **VERY SEVERE** limitations.

Using conventional resource analysis techniques the final road limitation rating was based on the most limiting factor. Accordingly, once all eight factors were derived the Final Limitation Factor (ROADFACT) was set to the worst (or most limiting) factor. This map represents the most limiting rating factor for each polygon.



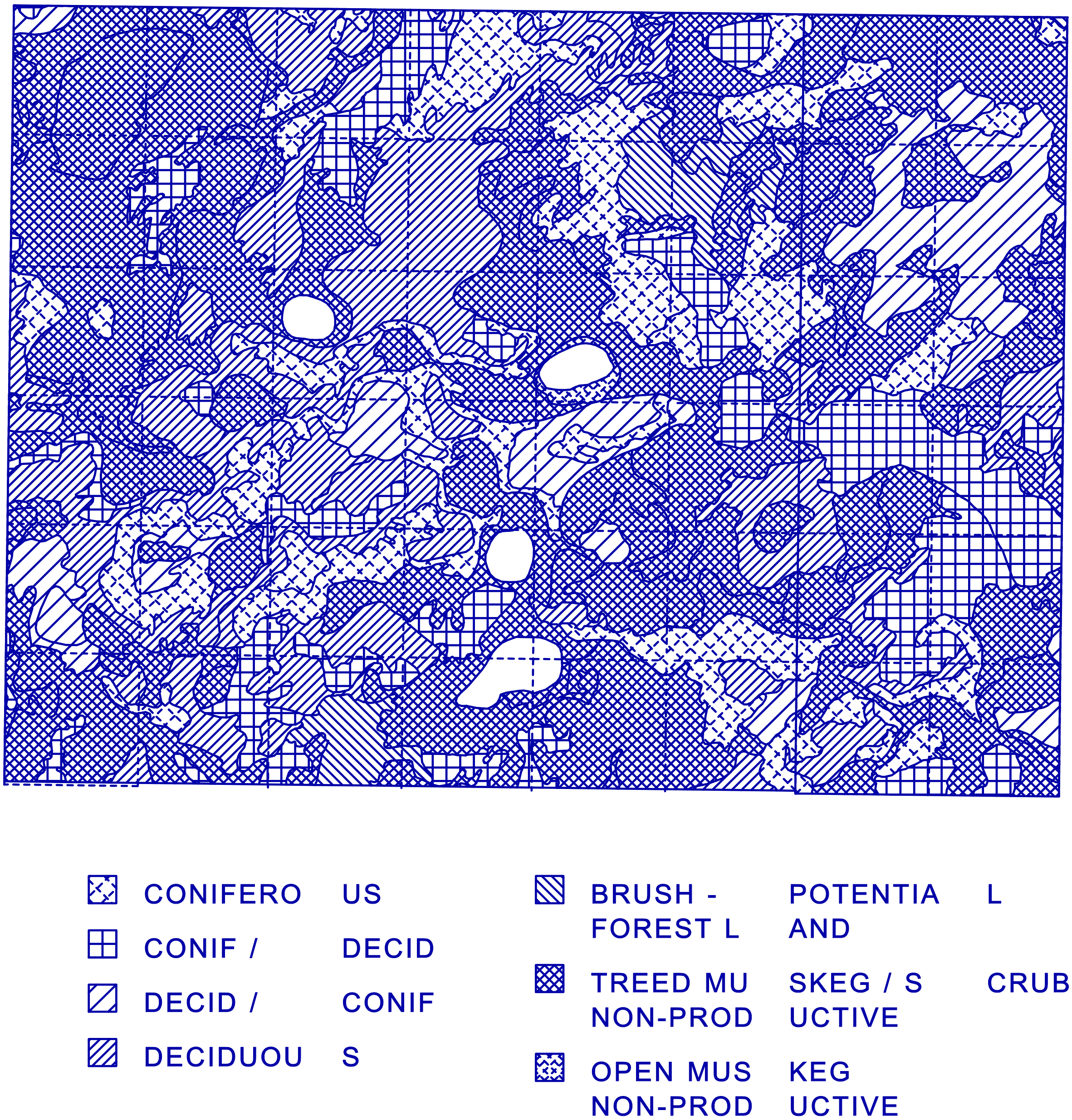
Example 1: Map 4 - This map represents the most limiting road location factor. Reclassification techniques were used to derive the final polygon values based on original and derived soil attributes.

The **third** set of maps (maps 5-7) include three reclassifications of the FOREST data layer. **Map 5** is a simple reclassification based on original commercialism codes.



Example 1: Map 5 - A reclassification of the FOREST layer based on the 'commercialism code' attribute.

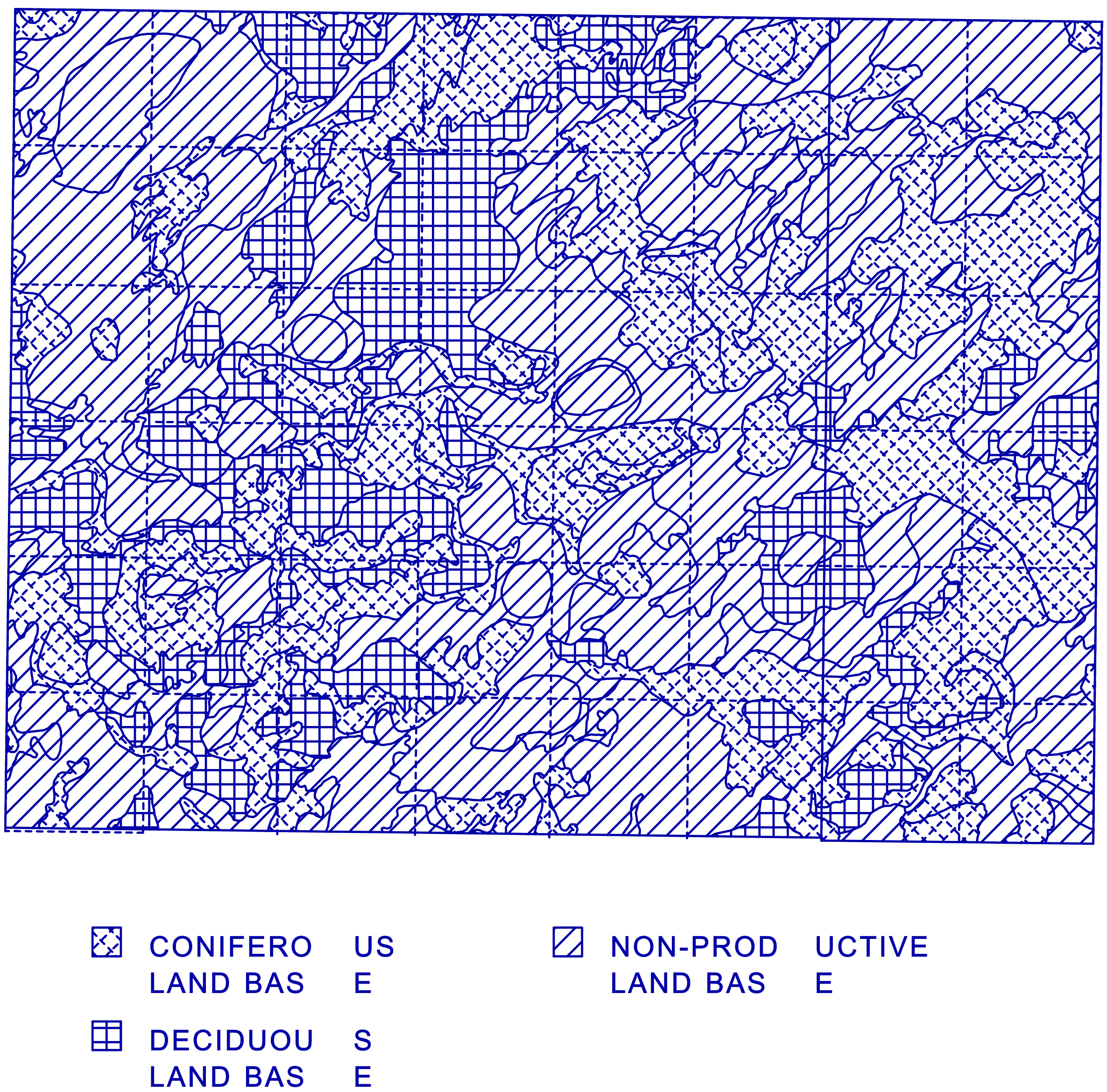
Map 6 is a simple reclassification based on original cover group classes. Note that the feature polygons are the same, only the data content has changed.



Example 1: Map 6 - A reclassification of the FOREST layer based on the 'cover group' attribute

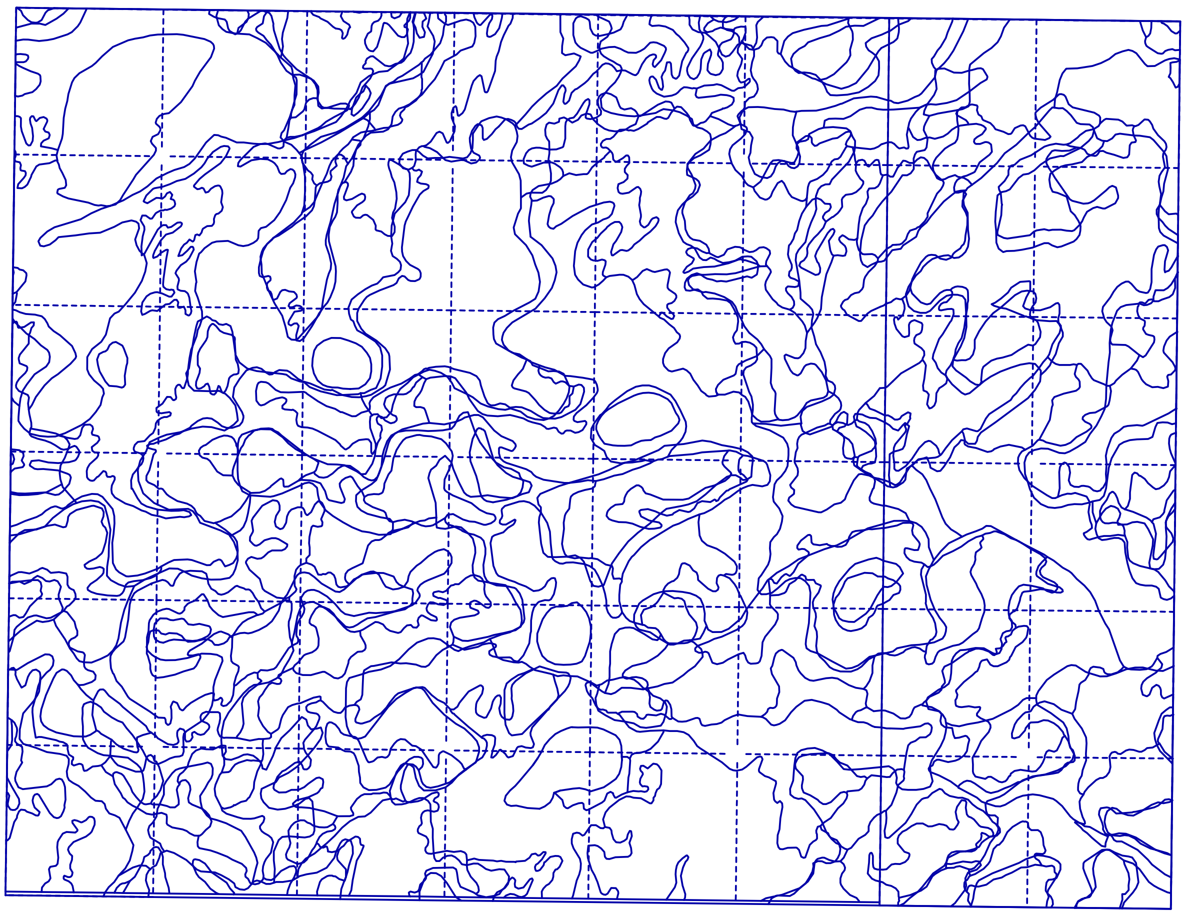
Map 7 is a simple reclassification where the cover group was generalized to create a coniferous / deciduous land base map. This merely involved aggregating cover group classes in the following manner :

> **CVGRP1** equals **1 to 3** then **Coniferous**
> **CVGRP1** equals **4** then **Deciduous**
> **CVGRP1** equals **5 to 7** then **Non-productive**.



Example 1: Map 7 - A reclassification of the FOREST layer to create a Land Base Map based on a generalization of the 'cover group' attribute

The **final set** of maps (maps 8-9) for the Resource Capabilities example involves a topological overlay of the SOIL and the FOREST. **Map 8** represents the combined data layer created by overlaying the two. Note that a total of **901** polygons were constructed by overlaying the original SOIL (71 polygons) and the original FOREST layers (187 polygons). This is representative of the increased complexity inherent in overlaying two data layers. Accordingly, the possibility of significantly increasing the proportion of error also exists.



Example 1: Map 8 - The new polygon data layer constructed by overlaying the SOIL layer with the FOREST layer. This layer contains 901 polygons. Note the existence of numerous sliver polygons indicating an accuracy and/or scale mismatch between the original data sets.

The overlay process constructs a new topological data layer by combining the spatial data from both original layers (SOIL and FOREST). During the overlay process the attribute tables are also *joined* together to form a new single table that will *link* to the new polygon layer. The schema of the new table is simply created by appending both existing tables together. The new schema is presented below.

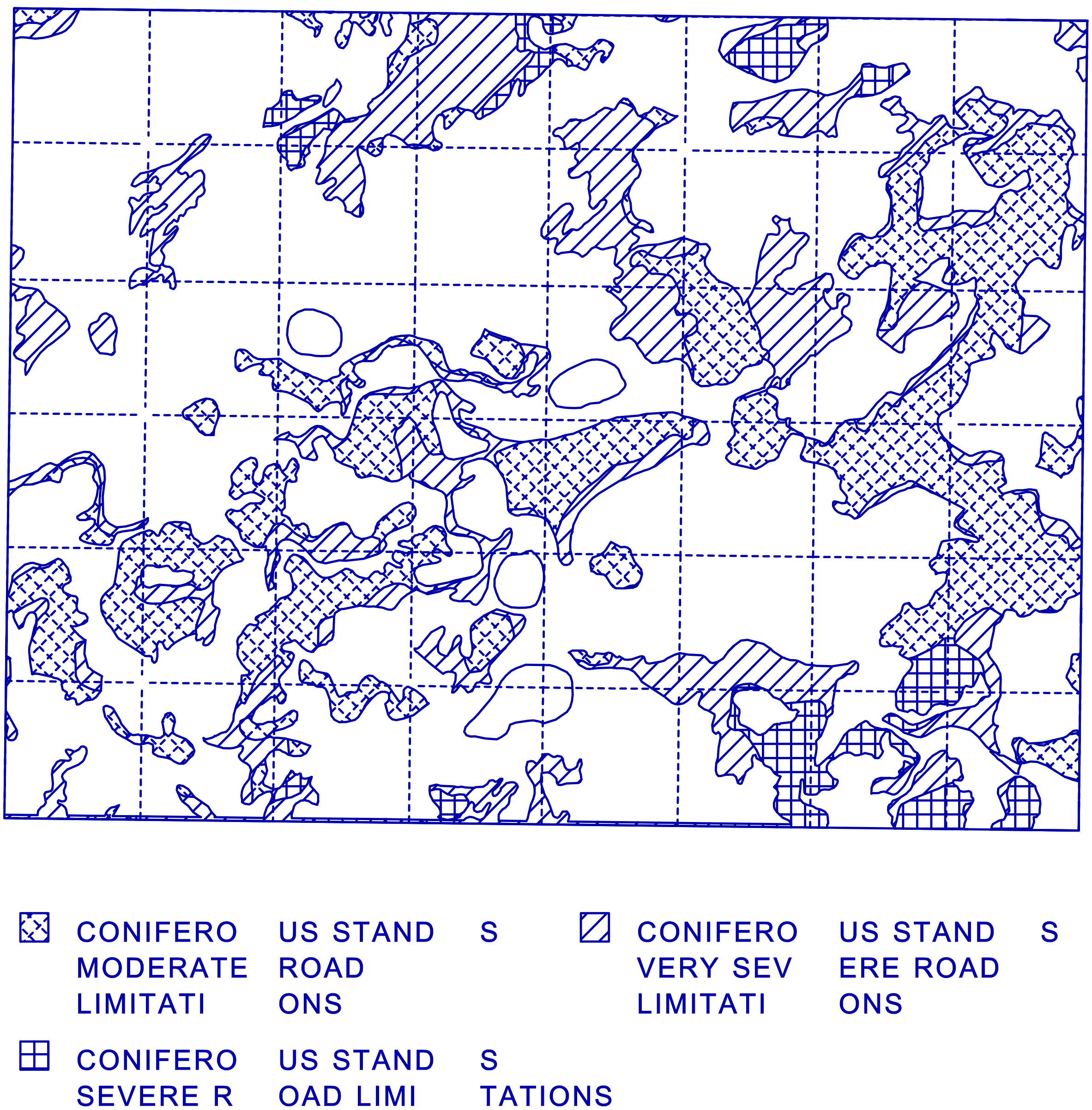
Attribute List - Overlaid SOIL and FOREST

This list describes the schema of the FOREST attribute table. The AREA, PERIMETER and SOILFOREST-ID columns are calculated automatically by the GIS software during the topological building process, e.g. in constructing clean topological polygons. The SOILFOREST-ID represents the unique identifier assigned to each new polygon automatically by the GIS software. The original polygon ids for SOIL and FOREST are maintained in the SOIL-ID and FOREST-ID columns.

Field	*Column Name*	*Description*
(New topological derived attributes)		
1	AREA	Area of polygon
2	PERIMETER	Perimeter of polygon
3	SOILFOREST-ID	Unique identifier (label)
(Original SOIL attributes)		
4	SOIL-ID	Unique identifier (label)
5	PARMAT	Parent material
6	SURFEX	Surface expression
7	SLOPE	Slope
8	SURFTEXT	Surface soil texture
9	UNDERTEXT	Underlying soil texture
10	SOIL	Soil classification
11	DRAIN	Drainage class
12	SERIES	Soil series
13	SUBREG	Physiographic subregion
(Derived SOIL attributes during reclassification)		
14	**ROADDRAIN**	**Drainage limitation for road location**
15	**ROADSLOPE**	**Slope limitation for road location**
16	**ROADSHRINK**	**Shrink-swell limitation for road location**
17	**ROADAASHO**	**AASHO index limitation for road location**
18	**ROADFROST**	**Frost heave limitation for road location**
19	**ROADROCK**	**Bedrock limitation for road location**
20	**ROADSTONE**	**Stoniness limitation for road location**
21	**ROADLIM**	**Final limitation rating for road location**
22	**ROADFACT**	**Limitation factor for road location**

Field	*Column Name*	*Description*
(Original FOREST attributes)		
23	FOREST-ID	Unique identifier (label)
24	FOREST	Forest code
25	UNIT	Management unit
26	MANAGER	Manager
27	TWP	Township number
28	RGE	Range number
29	MER	Meridian
30	STANDNO	Stand number
31	STANDAL	Stand number alphanumeric code
32	STORY	Map story - under or over
33	DENSITY	Crown density code
34	HEIGHT	Height code
35	SP1	Species One code
36	SP2	Species Two code
37	SP3	Species Three code
38	SP4	Species Four code
39	COM	Commercialism code
40	VSR	Volume Sampling region (VSR)
41	CSP1	Coded species code one
42	CSP2	Coded species code two
43	CSP3	Coded species code three
44	CSP4	Coded species code four
45	CVGRP1	Cover group one
46	CVGRP2	Cover group two
47	SITE	Site index code
48	SLOPE	Slope code
49	PLANZONE	Planning zone
50	G/W	G/W zone code
51	OCEAN	Ocean drainage
52	BASIN	Drainage basin
53	WATERSHED	Watershed
54	SUBWATER	Sub watershed
55	ORIGIN	Date of origin
56	CON15	Live coniferous volume 15+/11 cm
57	DEC15	Live deciduous volume 15+/11
58	CON20	Live coniferous volume 19+/13
59	DEC20	Live deciduous volume 19+/13
60	CON25	Live coniferous volume 25+/15
61	DEC25	Live deciduous volume 15+/15
62	ORIGMER	Origin class / merchantability

Map 9 represents a conceptual overlay of the FOREST derived *Forest Land Base* map and the SOIL derived *Soil Limitations* map. This overlay produced a *Road Limitations for Coniferous Land Base* map. In fact, while this process can be conceptually thought of as an overlay between the two products, it was actually derived from the overlay of the SOIL and the FOREST. Given that the SOIL reclassifications were done prior to the overlay, a multiple reclassification using the FOREST *Forest Land Base* logic and the SOIL ROADLIM attribute of the *Soil Limitations for Road Location* logic, was undertaken to create a classified map representing this map.



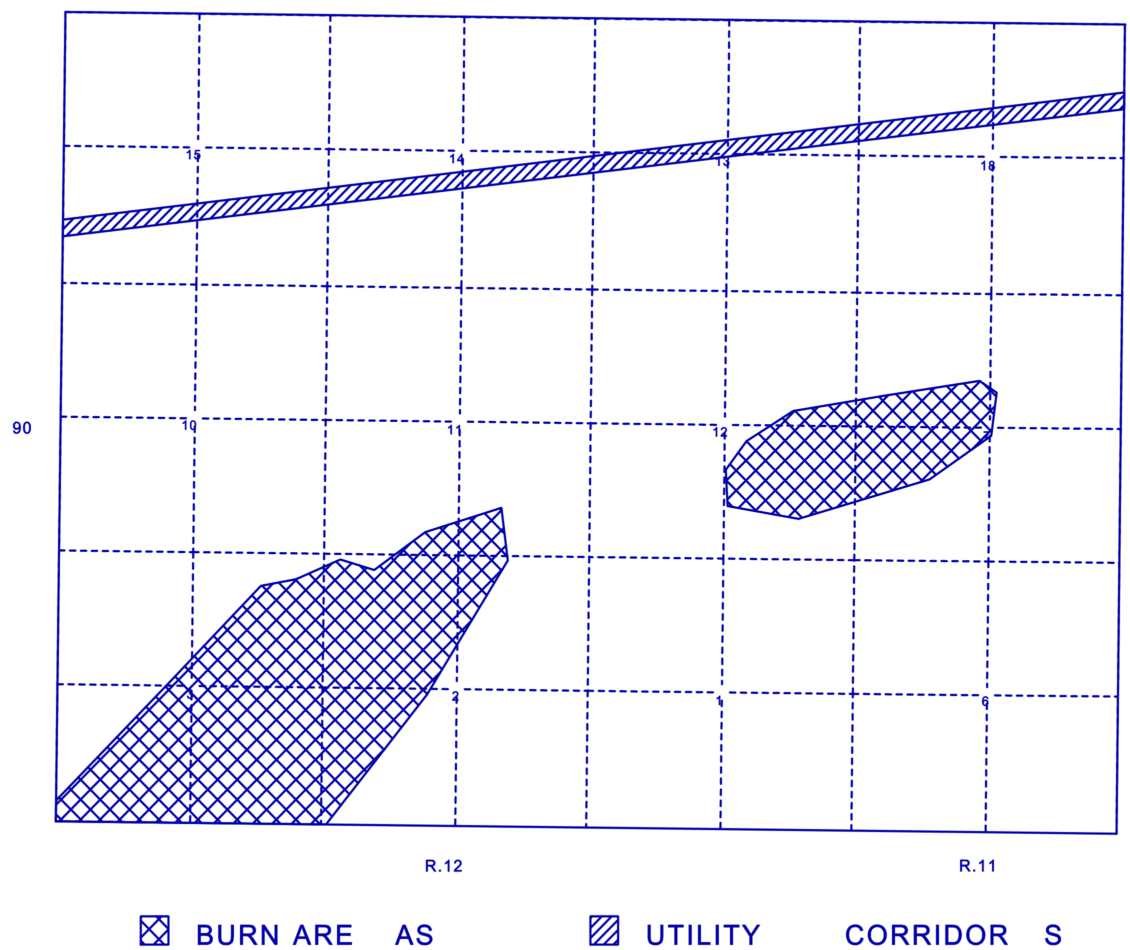
Example 1: Map 9 - A 'Road Limitations for Coniferous Land Base' map derived by reclassifying both selected SOIL and FOREST attributes after a topological overlay had taken place.

It should be noted that the distinction between reclassification and topological overlay varies considerably among GIS software. In some systems the final set of maps can be termed strict topological overlay, while others required a reclassification as well. The exact technique is based on the architecture of the GIS, especially the design of the functional interaction between the GIS query tools and the DBMS functions.

As well, the reclassification and overlay examples presented only reflect a vector data model approach. Reclassification and overlay are both supported using the raster data model using a different procedural logic. Vector data model approaches are traditionally done with forestry type data, however with the integration of both vector and raster processing tools, most spatial modeling and overlay analytics are migrating towards use of the raster data approaches.

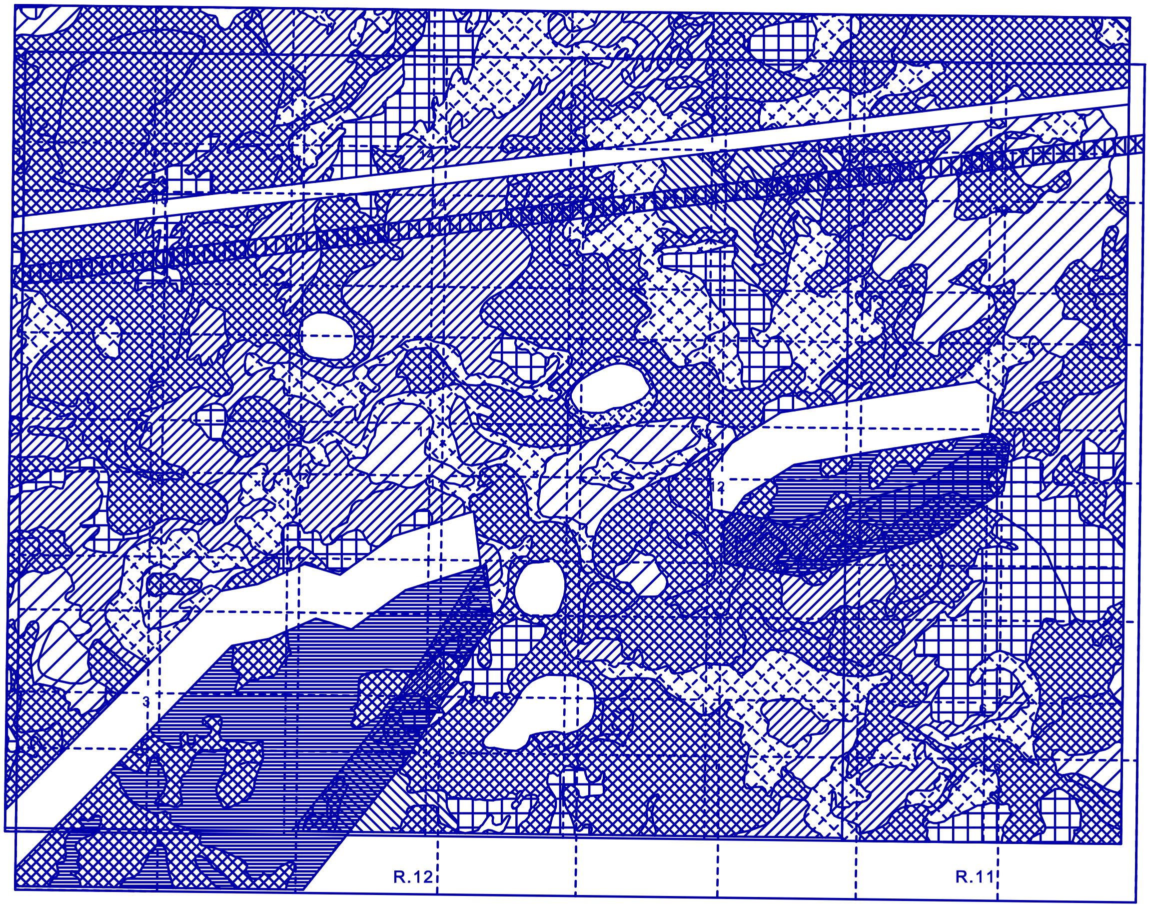
6.2 Example 2. Forest Inventory Updating

This sample application uses the forest inventory (FOREST) data layer and a forest depletion data layer. The FOREST layer contains **187** polygons. The DEPLETION layer contains **3** polygons. **Map 1** presents the depletion areas. The DEPLETION data layer consists of two burn areas and one right-of-way cut for a utility corridor.



Example 2: Map 1 - A sample depletion layer consisting of two burn areas and a utility

Map 2 represents the forest inventory layer topologically overlaid with the depletion data layer to create a revised inventory layer with holes. The holes represent the physical areas bounded by the depletion layer. In this regard a 'clipping' overlay function was applied to remove the depletion areas first from the existing FOREST layer. The new overlaid layer is presented classified by cover group.



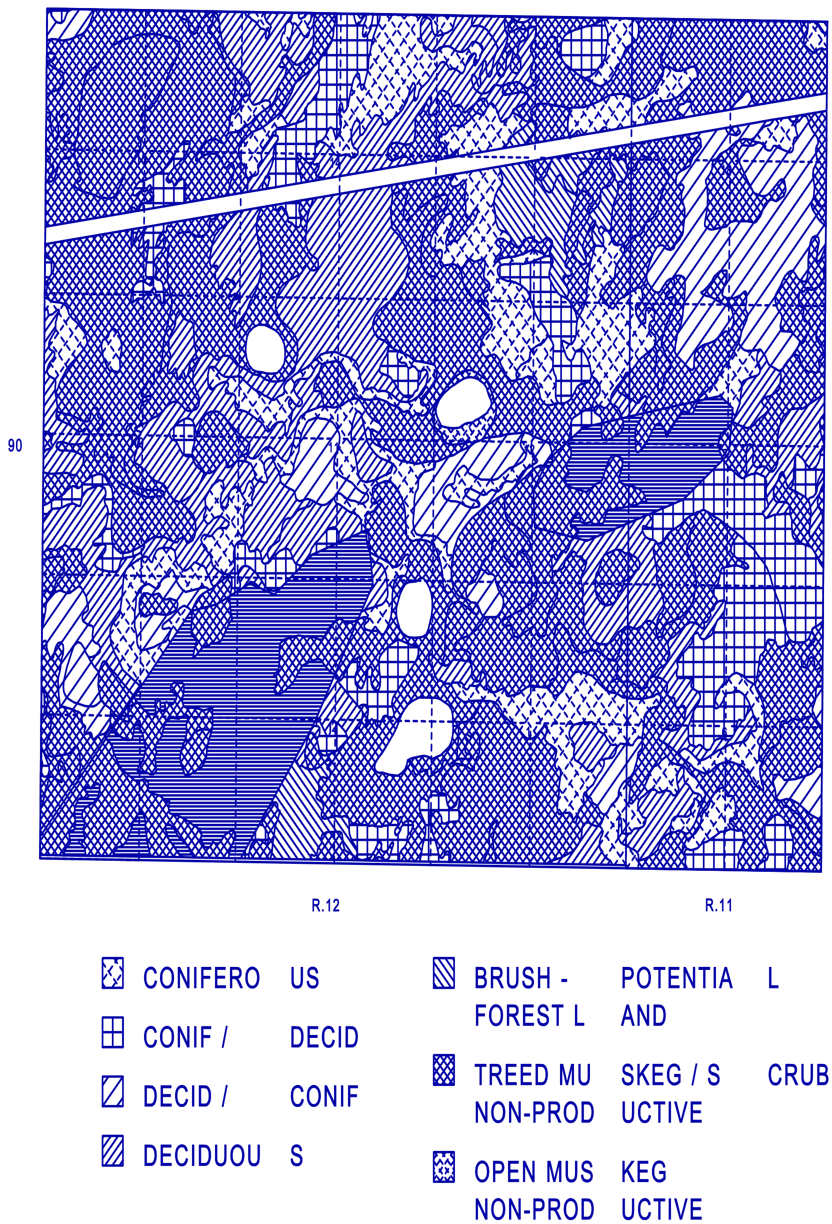
Example 2: Map 3 - This map represents another topological overlay between the FOREST and DEPLETION layers. In this case only forest cover within the depletion areas was clipped and classified by cover group to determine which types 'would' and 'would not' burn.

Example 2: Map 2 - This map represents a topological overlay undertaken with a 'clipping' function between the original FOREST layer and the DEPLETION layer. Depletion areas have been clipped out of the forest cover.

Map 3 represents the areas of depletion classified by forest cover group. This was achieved by topologically overlaying the original forest inventory layer and the depletion layer in a clipping fashion only retaining the inventory data within the depletion layer boundaries. The inventory data within the depletion boundaries was then classified as to their potential for burning. Boundaries for the burn areas were dissolved to create a homogeneous depletion data layer.

Map 4 represents a topological overlay of the revised inventory layer and the classified depletion layer to create an updated forest inventory data layer. Appropriate attributes for the modified inventory polygons were updated in the process as well. The final product represents a completely updated inventory where existing stands that would burn were replaced, and those not suitable in a burn scenario, eg. muskeg, were left unchanged.

Since whenever a topological overlay occurs attribute tables are also merged, reports as well as map can also be generated using conventional DBMS reporting capabilities. For this example two reports were also generated for the update process. Summary reports on the area and volume lost for all CONIFEROUS forest stands within the burned area and corridor area were created. Volumes are for the 15+/11 centimetre utilization standard.



Example 2: Map 4 - This map represents an updated forest inventory. The revised forest cover has been 'merged' with the reclassified depletion layer. Note that cover groups that

It is important to note that this example used topological overlay in a variety of ways. The previous maps each utilized a different Boolean operator to achieve the results. For example, creation of the revised inventory layer with holes makes use of the NOT operator, eg. FOREST .NOT. DEPLETION. Creation of the classified depletion layer makes use of the AND operator followed by a dissolving of boundaries, eg. FOREST .AND. DEPLETION. The final updated forest inventory layer makes use of the OR operator, eg. FOREST .OR. DEPLETION.

Example 2. Forest Inventory Updating
Report on BURNED AREAS - Volume and Area Depletion
(Coniferous Landbase)

STND NUM	COV	D	SP	CO	YR AGE	AREA	CUM. AREA	VOLUME	CUMUL. VOLUME
68	1	A	SW	R	1860	6.29	6.29	584.45	584.45
72	2	B	SW	L	1860	24.40	30.69	2896.45	3,480.92
76	2	A	SW	L	1860	10.72	41.41	1510.27	4,991.19
84	3	C	AS	U	1940	0.20	41.61	5.02	4,996.21
85	2	B	SW	L	1860	0.93	42.54	112.32	5,108.53
101	1	B	SW	L	1840	26.89	69.42	4011.56	9,120.09
106	1	A	SW	U	1840	8.09	77.51	803.75	9,923.84
113	3	C	AS	U	1940	0.00	77.51	0.04	9,923.88
115	1	B	SW	L	1840	7.11	84.62	1059.67	10,983.55
121	2	B	SB	U	1940	24.37	108.99	638.43	11,621.98
125	2	C	SW	R	1940	6.84	115.82	1282.41	12,904.39
135	1	C	PS	U	1940	5.97	121.79	319.82	13,224.21
143	2	C	SB	U	1940	4.24	126.03	138.67	13,362.88
153	2	B	SB	U	1940	1.79	127.82	47.17	13,410.06
154	1	D	SB	U	1840	0.27	**128.09**	13.30	**13,423.35**

Total Area lost = **128.09** (HA.)
Total Volume lost = **13,423.35** (CU.M)

Report on CORRIDOR AREAS - Volume and Area Depletion
(Coniferous Landbase)

STND NUM	COV	D	SP	CO	YR AGE	AREA	CUM. AREA	VOLUME	CUMUL. VOLUME
5	3	A	AS	U	1860	3.11	3.11	165.86	165.86
10	1	C	SB	U	1940	0.17	3.29	10.40	176.27
12	3	A	AS	U	1860	1.19	**4.48**	63.63	**239.90**

Total Area lost = **4.48** (HA.)
Total Volume lost = **239.90** (CU.M)

6.3 Example 3. Merchantability and DEM Modelling

The **third** example integrates Digital Elevation Model (DEM) data and forest inventory data. The study area differs from the first two applications. It represents one township in an area of extreme relief. It is an excellent example for illustrating DEM modelling functions.

Map 1 illustrates the distribution of elevation data points. A total of 16,443 irregular points are represented. Four different kinds of data points were present in the DEM. The DEM overlaps the study area by at least 200 metres. This is required to properly allow the GIS to interpolate values on the edge of the study area. The DEM data was photogrammetrically captured from 1:60,000 scale aerial photography and compiled for 1:20,000 mapping standards. This DEM was loaded into the GIS and interpolated into a 50 metre raster grid for the study area. All DEM modelling was done from the interpolated grid.

Map two illustrates the forest inventory polygons for the study area. There are **474** stands. **Map three** illustrates slope classes based on the 45 % standard for logging. The slope classes were defined to identify optimum slope areas for logging potential. Note the raster representation of the interpolation. This reflects that level of data interpolation for the DEM grid. These maps represent sample products that can be generated from the DEM data. Elevation ranges, slope, and aspect are commonly regarded as *first derivative* DEM products. A wide variety of secondary, or **second derivative**, products can also be generated from DEM data. Use of any DEM products is clearly dependant on the aim and requirements of the GIS project.

Map four represents merchantable coniferous stands classed by age. Boundaries were *dissolved* between polygons with the same rating.

Map five represents a topological overlay between the merchantable stands and the slope. A final rating scheme was devised based on slope class and merchantability class. For example, to determine three classes a rating matrix was defined. The model can be represented as :

if **SLOPE-CLASS** = **1 OR 2** and **MERCH-CLASS** = **1** then **MS-RATING** = '**H**'
if **SLOPE-CLASS** = **2** and **MERCH-CLASS** = **2** then **MS-RATING** = '**H**'

if **SLOPE-CLASS** = **1** and **MERCH-CLASS** = **2 OR 3** then **MS-RATING** = '**M**'
if **SLOPE-CLASS** = **2** and **MERCH-CLASS** = **3** then **MS-RATING** = '**M**'
if **SLOPE-CLASS** = **3** and **MERCH-CLASS** = **1 OR 2** then **MS-RATING** = '**M**'

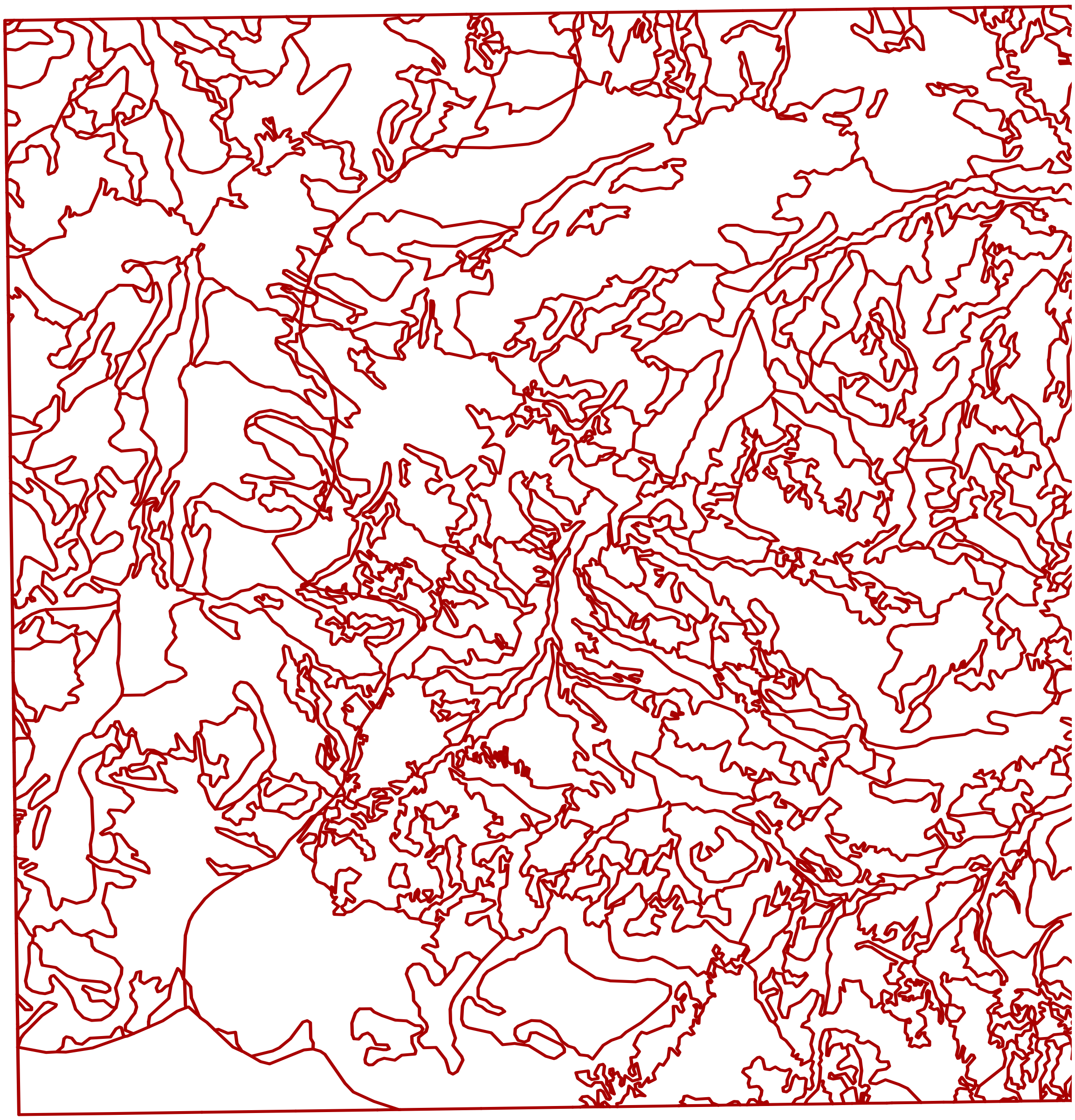
All other combinations **MS-RATING** = '**L**'

The model logic can also be represented in matrix form where slope class is represented by columns and merchantability class by rows :

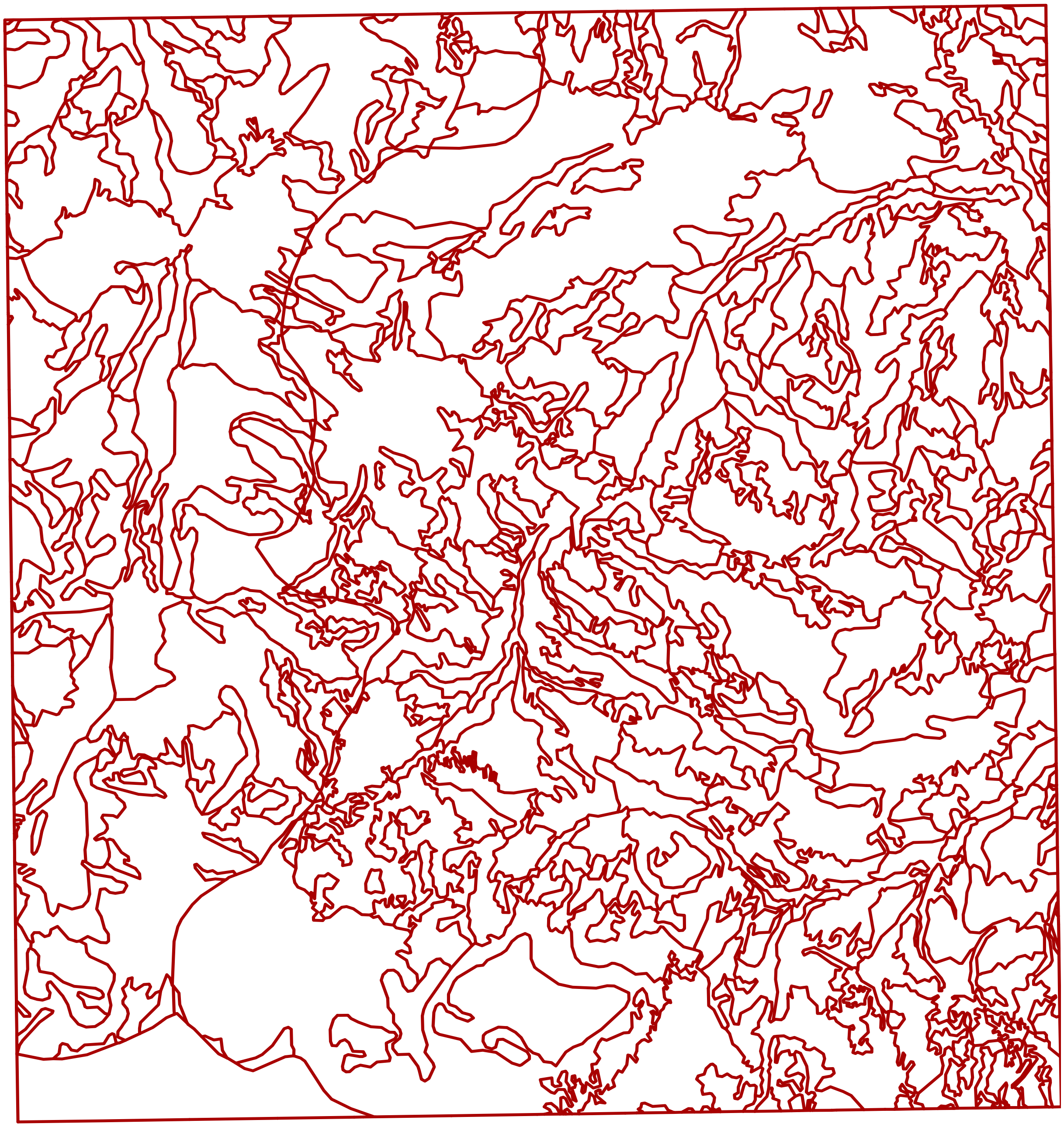
S L O P E   C L A S S

	1	**2**	**3**	**4**
1	**H**	**H**	M	L
2	**H**	**H**	M	L
3	M	M	L	L
4	L	L	L	L

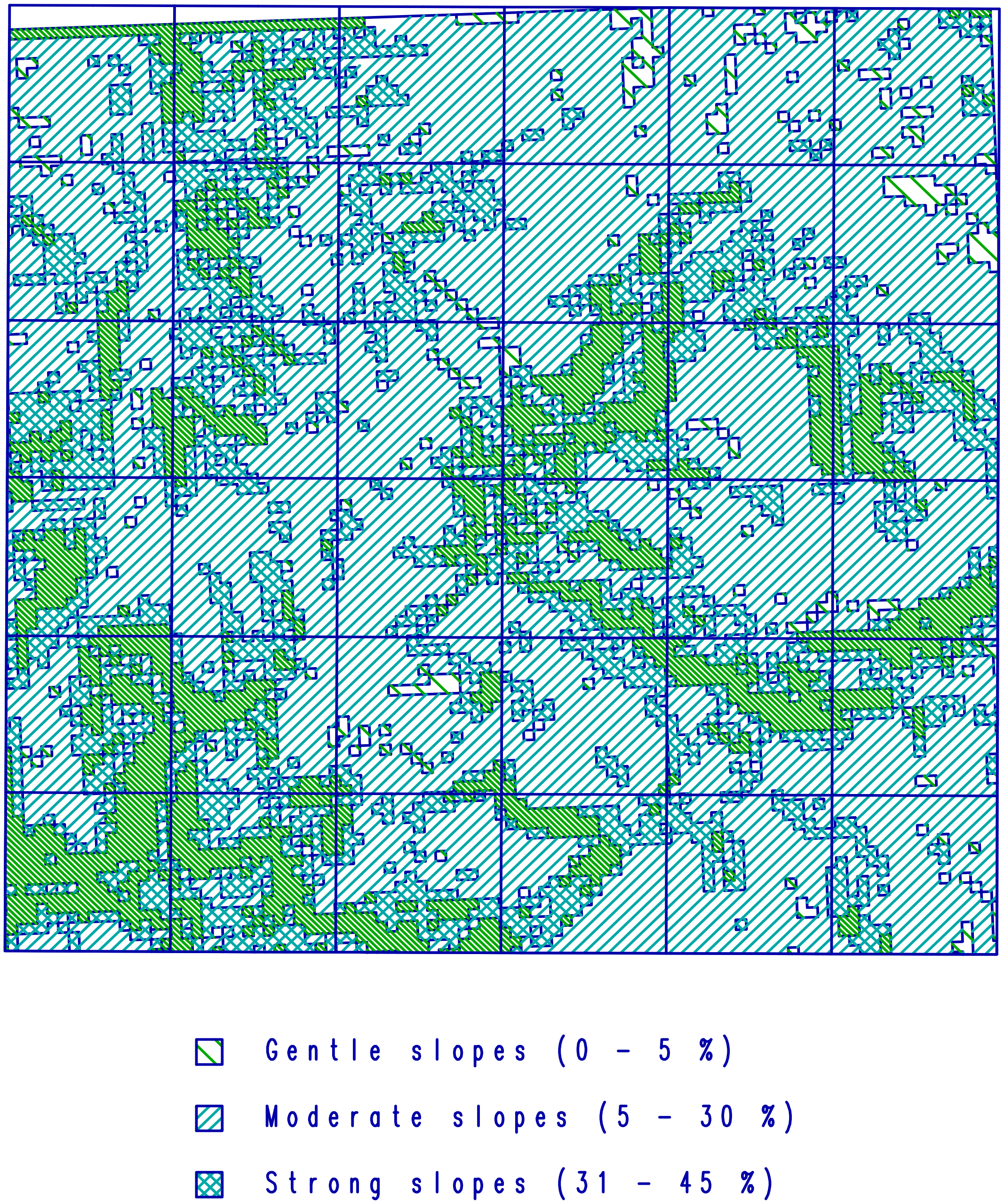
Note that in the maps polygon boundaries were dissolved between areas of common rating.



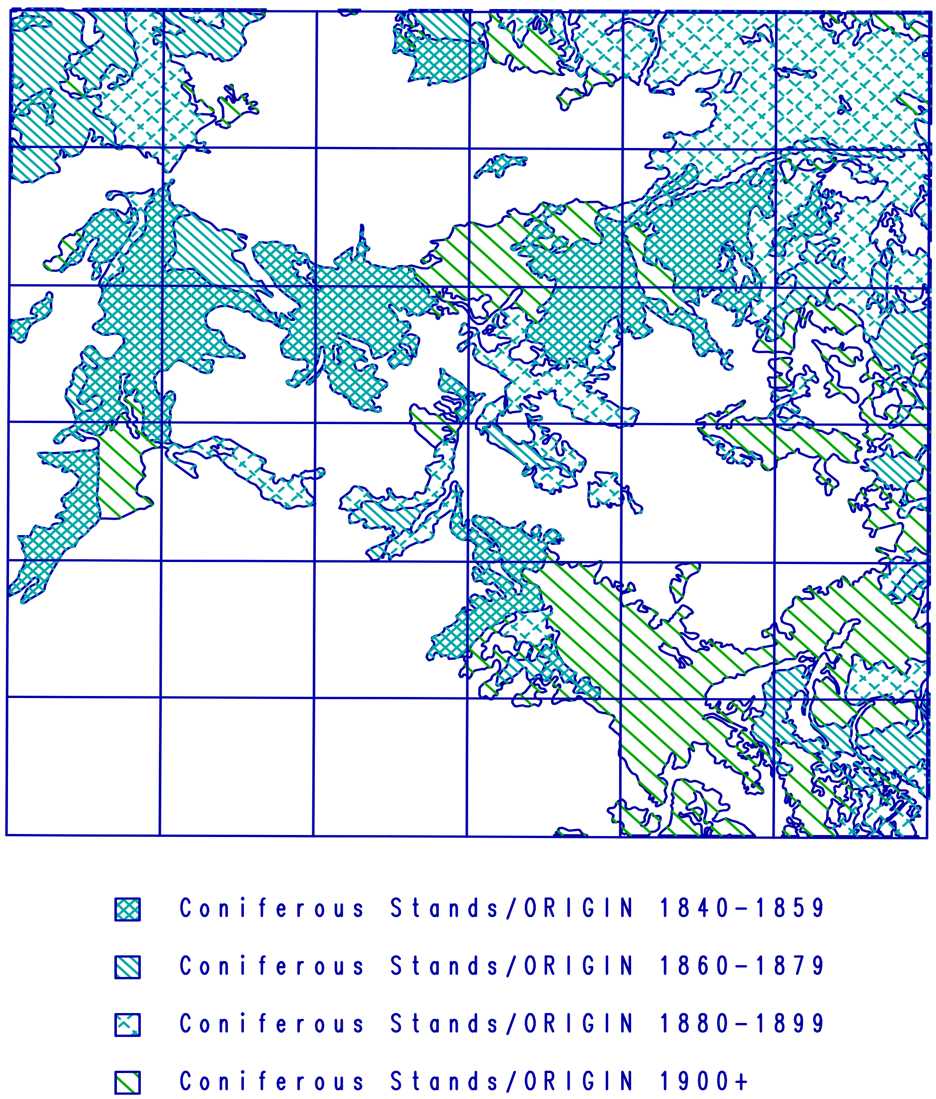
Example 3: Map 1 - Digital Elevation Model points.



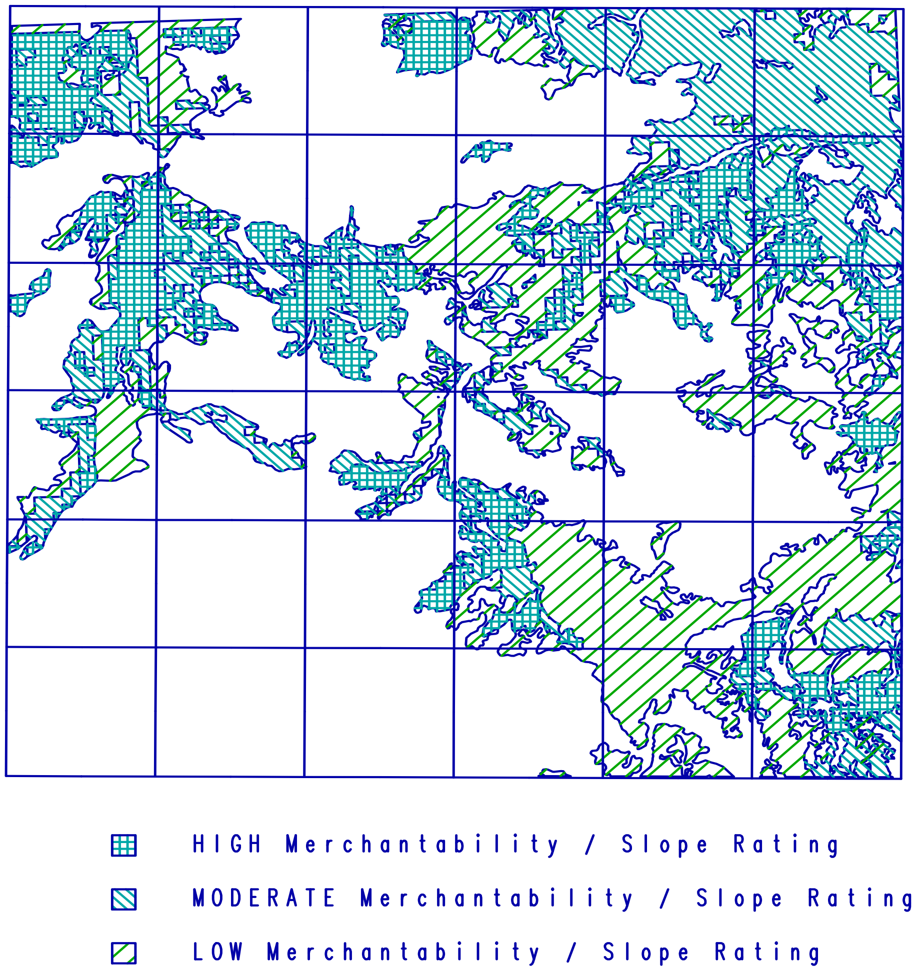
Example 3: Map 2 - Forest Inventory Polygons



Example 3: Map 3 - Slope Classes



Example 3: Map 4 - Merchantable Coniferous Stands by Age



Example 3: Map 5 - Merchantability by Slope

7.0 Data Output and Display

This chapter reviews typical data output capabilities and requirements. The focus is on describing map output, map composition, and tabular reporting requirements.

The fourth and final subsystem for in a GIS is the data output and display subsystem. This subsystem allows the user to generate graphic displays, normally maps, and tabular reports. Data output is the operation of presenting the results of data analysis in a form that is understandable to a user, or in a form that allows data transfer to another computer system.[38] This discussion of data output focuses on user oriented output.

7.1 Graphic Display

Graphic display is the primary output mechanism of GIS. Most often maps are the graphic result. Graphic display occurs in several modes. Usually, graphic products are displayed on a video display terminal immediately following the generation of results. Like the DBMS's, some GIS vendors use a proprietary graphics software for the display of output. Some others use graphic primitives such as the *Graphic Kernel System (GKS)* to provide a more flexible interface to graphics hardware. The GKS provides the programmer with a subroutine library of two-dimensional graphics primitives that can be accessed through a consistent interface in high-level computer language such as FORTRAN.[39] Standards such as GKS can operate in a vector or raster mode.

Graphic displays can also be generated for *plotting* on a hardcopy device. Hardcopy devices can include pen plotters, raster plotters, e.g. electrostatic, ink jet, thermal wax, and optical film writers. Each type of hardware device can vary greatly in complexity and quality of output. Accordingly, they also range greatly in price. Most GIS software provides interfaces to create *plot files* in a variety of plotter formats. The plotting industry has standardized their plotting subroutines so most GIS vendors can easily adapt their software to the specific plotting device of a purchaser. A wide variety of plotters are usually supported.

While maps are the most common graphic output product, some applications require the capability for generating graphic output such as histograms, scattergrams, profiles, and charts. Some GIS packages have neglected this capability in the past, however most realize the need for such statistical output options and are attempting to incorporate these functions in their new releases. Many vendors are solving this lack of functionality by providing format conversion utilities so users can easily transfer data from the GIS into third-party statistical software packages.

---

[38] Burrough, 1986.

[39] Burrough, 1986.

7.2 Map Composition

As GIS software is becoming used more for decision support there is a need for cartographic quality output products. The ability to *compose* maps using standard cartographic principles is becoming increasingly important. Most GIS software achieves map composition by allowing users to build *map templates*. Once these map templates are built they can be used to generate plot files independent of the data set being utilized. Map composition capabilities typically include the capability to define *viewports* or windows into the data set. Often viewports are defined by map scale, e.g. 1:15,000. Viewports can also be utilized for tabular reports, perspective views, and statistical graphic displays, e.g. histograms. The primary map composition requirements include defining :

- color;
- feature symbology for point and lines, e.g. line width, line type, line pattern, point symbols, e.g. highway symbols;
- annotation, e.g. text labelling and descriptions;
- map scale bar and source;
- legend; and
- polygon shading symbology, e.g. typically cross-hatching or color fill.

As with many of functions these capabilities vary greatly across the different commercial GIS software. Users will generally find that these capabilities are well developed in CAD and mapping software but often leave much to be desired with full featured GIS software.

The *user interface* for map composition also varies considerably between GIS software offerings. Users are generally provided with a *command language* or *menu driven interface*. Initial user normally prefer the *point-and-shoot* style of the mouse driven menu interface while more experienced users often prefer the command driven interface. Graphics definitions and map composition is usually easier with the menu driven approach.

7.3 Tabular Reports

The second type of output that is commonly required by GIS users is the tabular or database report. *Tabular reports* are database listings of the results of GIS analysis matching the graphic display. Frequently, a user will generate a product based on some defined analysis process. The map is usually the primary product. However, users often require some statistics to summarize the result. Resource applications are often concerned with total area and volumes. The tabular report is usually generated from the DBMS. Report templates can be established to formalize a reporting format for a particular production process. The report generated in the example applications were generated from user defined templates. With this approach any data set can be input for reporting if the attribute columns that are required exist for the data layer.

The *statistical report* implies a statistical analysis of the results. Some applications require the use of standard statistical methods for the analysis of processing results. The generation of means, modes, averages, totals, regression analysis, correlation analysis, etc. are good examples. Some DBMS's within the GIS do provide a number of these minor functions. However, most do not provide full statistical analysis techniques matching any of the standard approaches found in geography. Often, the user must *export* data from the GIS into another software package, e.g. SAS, to have this capability. This is usually not a concern as most application users are familiar with these packages well before their GIS software, and the export process is quite satisfactory. Remember that the GIS is a specialized tool and often lacks many of the standard information processing options found in non-spatial oriented systems.

8.0 Implementation Strategies and Issues

This chapter presents an overview of implementation issues and requirements. The focus is on identifying implementation planning issues and strategies that must be addressed for a successful GIS implementation. This chapter will be of most interest to managers.

8.1 Current Options and Software Assessment

Perhaps the first question asked by anyone when discovering GIS is *what are the current options available ?*. This question is often asked as directly as *what is the best GIS ?*. Quite simply, there is no best GIS. A wide variety of GIS software offerings exist in the commercial market place. Commercial surveys often are a good starting point in the assessment of GIS software. The number of GIS software offerings is approximately 10 if one eliminates the following :

- the university based research software, which tends to lack full integration and usually has narrow channels of functionality;
- the CAD vendors, who like to use GIS jargon but often cannot provide full featured functionality; and
- the consulting firms, that will provide or customize selected modules for a GIS but lack a complete product.

One of the problems in evaluating the functionality of GIS software is the bias one gets from using one system or another. Comparing similar functions between systems is often confusing. Like any software, ultimately some do particular tasks better than others, and also some lack functionality compared to others.

> **Due mostly to this diverse range of different architectures and the complex nature of spatial analysis no standard evaluation technique or method has been established to date.**

Any GIS should be evaluated strictly in terms of the potential user's needs and requirements in consideration of their work procedures, production requirements, and organizational context ! The experienced GIS consultant can play a large and valuable role in the assessment process.

A current accepted approach to selecting the appropriate GIS involves establishing a benchmark utilizing real data that best represents the normal workflow and processes employed in your organization.

The identification of potential needs and requirements is essential in developing a proper benchmark with which to evaluate GIS software packages. A formalized user need analysis is absolutely critical to the successful implementation of GIS technology.

Development of the benchmark should include a consideration of other roles within your organization that may require integration with the GIS technology. A logical and systematic approach as such is consistent with existing information systems (IS) planning methodologies and will ultimately provide a mechanism for a successful evaluation process.

8.2 Practical Considerations

A host of practical considerations in the acquisition of GIS technology exist. The major concerns are identified below. This review is oriented towards the potential GIS user not the existing one.

8.2.1 Data Storage Volumes

Perhaps the primary concern when using a GIS is the data volumes that will be generated from your GIS. Experience has shown that the data required to complete a GIS project is often far beyond initial estimates. Even for small local applications the mass of data usually exceeds expectations. A strict product description and systematic project plan will help to keep data volumes down. The definition of explicit project objectives and deliverables within a required timeframe will also help to limit the generation of extraneous data. Extraneous, but yet often useful data, is often generated due to the ease of modelling within the GIS environment and the ad hoc nature of the DBMS querying environment. Beware data proliferates !

Data volumes have been of specific concern in the micro-computer based GIS environment. Most micro based GIS users are limited in the amount of storage available to them. Even with the recent technological developments in hardware, dealing with large data sets in the micro environment is often cumbersome and sometimes impossible. The implications of processing times for large data sets should also be a concern. Shoppers should strategically plan for the necessary hardware to support any GIS product that is purchased. Vendors are quite valuable in helping users establish the appropriate system configuration to support the data volumes for their particular applications.

8.2.2 Software Processing Limitations

A consideration when assessing GIS software is the constraints or limitations of the software. Generally, large data volumes mean a large number of features. Potential buyers of GIS should investigate the processing limitations and constraints of the GIS software they are interested in. Most systems have distinct limitations with respect to the number of features supported during compilation and analysis.

Some of the more obvious limitations include:

Raster Systems

- the maximum number of rows and columns in a data layer;
- the minimum and maximum resolution size for a grid-cell;
- any specific algorithm restrictions, e.g. buffering limits, generalization limits;
- raster/vector conversion restrictions.

Vector Systems

- the maximum number of nodes in a data layers;
- the maximum number of arcs in a data layer;
- the maximum number of polygons in a data layer;
- the maximum number of arcs that can terminated at a node;
- the maximum number of arcs, e.g. line segments, in a polygon;
- the maximum coordinates pairs per arc; and
- any spatial indexing restrictions on the merging of adjacent data.

General

- are algorithms documented, e.g. standards, constants, default values, etc.;
- the maximum number of overlays permitted at one time;
- types of plotters, graphics adapters, printers supported;
- training requirements and options;
- data conversion interfaces available;
- application programming interface (API) available;
- macro language available;
- command and/or menu driven interface;
- maximum number of attributes for a data layer.

8.3 Justification and Expectations

GIS is a long term investment that matures over time. The turnaround for results may be longer term than initially expected. Quite simply, GIS has a steep learning curve. The realization of positive results and benefits will be not achieved overnight.

Both initial investment funding and continued financial support are major determinants in the success or failure of a GIS.[40]

[40] Burrough, 1986.

Most often the justification and acquisition of a GIS centers on technical issues of computer hardware and software, functional requirements, and performance standards. But experience has shown that, as important as these issues may be, they are not the ones that in the end determine whether a GIS implementation will succeed or not.[41]

Even though the proper assessment of an appropriate GIS product requires a good understanding of user's needs, most often systems are acquired based on less than complete and biased evaluations. Nonetheless, even with the GIS in hand a properly structured and systematic implementation plan is required for a successful operation. Generally, a GIS implementation plan must address the following technical, financial, and institutional considerations[42] :

- system acquisition tactics and costs;
- data requirements and costs;
- database design;
- initial data loading requirements and costs;
- system installation tactics, timetable, and costs;
- system life cycle and replacement costs;
- day-to-day operating procedures and costs;
- staffing requirements and costs;
- user training and costs; and
- application development and costs.

Potential GIS buyers should be aware of the necessary investment required in hardware, software, training, supplies, and staffing. The cost of establishing a successful GIS operation is substantial. However, with realistic expectations and support the development of GIS within an organization that manipulates geographic data will almost certainly prove beneficial.

Certain considerations of data longevity, data capture, personnel hiring, etc. are the practical concerns of GIS implementation. The longer term implications, such as hardware/software maintenance and replacement, should also be considered. The acquisition of GIS technology should not be done without seriously considering the way in which GIS will interact with the rest of the organization.

It is simply not enough to purchase a computer, a plotter,

[41] Aronoff, 1989.

[42] Jordan, 1989.

a display device, and some software and to put it into a corner with some enthusiastic persons and then expect immediate returns. A serious commitment to GIS implies a major impact on the whole organization.[43]

8.4 Implementation Issues

The mere presence of an implementation plan does not guarantee success. Most organizations do not have sufficient staff to cope with the commitment and extra work required when introducing a GIS to existing operations. GIS implementation must also consider all technology transfer processes.

8.4.1 Common Pitfalls

Several pitfalls exist that most often contribute to the failure of a GIS implementation strategy. These are identified below[44] :

- **Failure to identify and involve all users**

  Users in an operational GIS environment consist of operations, management, and policy levels of the organization. All three levels should be considered when identifying the needs of your users.

- **Failure to match GIS capability and needs**.

  A wide spectrum of GIS hardware and software choices currently exist. The buyer is presented with a significant challenge making the right choice. Remember, the right choice will be the GIS that provides the needed performance no more, no less for the minimum investment.[45] The success of a GIS implementation is particularly sensitive to the right hardware and software choices !

- **Failure to identify total costs**.

  The GIS acquisition cost is relatively easy to identify. However, it will represent a very small fraction of the total cost of implementing a GIS. Ongoing costs are substantial and include hardware and software maintenance, staffing, system administration, initial data loading, data updating, custom programming, and consulting fees.

- **Failure to conduct a pilot study**

[43] Burrough, 1986.

[44] after Jordan, 1988.

[45] Burrough, 1986.

The GIS implementation plan concerns itself with the many technical and administrative issues and their related cost impacts. Three of the most crucial issues, are database design, data loading and maintenance, and day-to-day operations. The pilot study will allow you to gather detailed observations, provided it is properly designed, to allow you to effectively estimate the operational requirements.

- **Giving the GIS implementation responsibility to the EDP Department**.

  Because of the distinct differences of the GIS from conventional EDP systems, the GIS implementation team is best staffed by non-data processing types. The specialized skills of the 'GIS analyst' are required at this stage. Reliance on conventional EDP personnel who lack these skills will ensure failure.

- **Failure to consider technology transfer.**

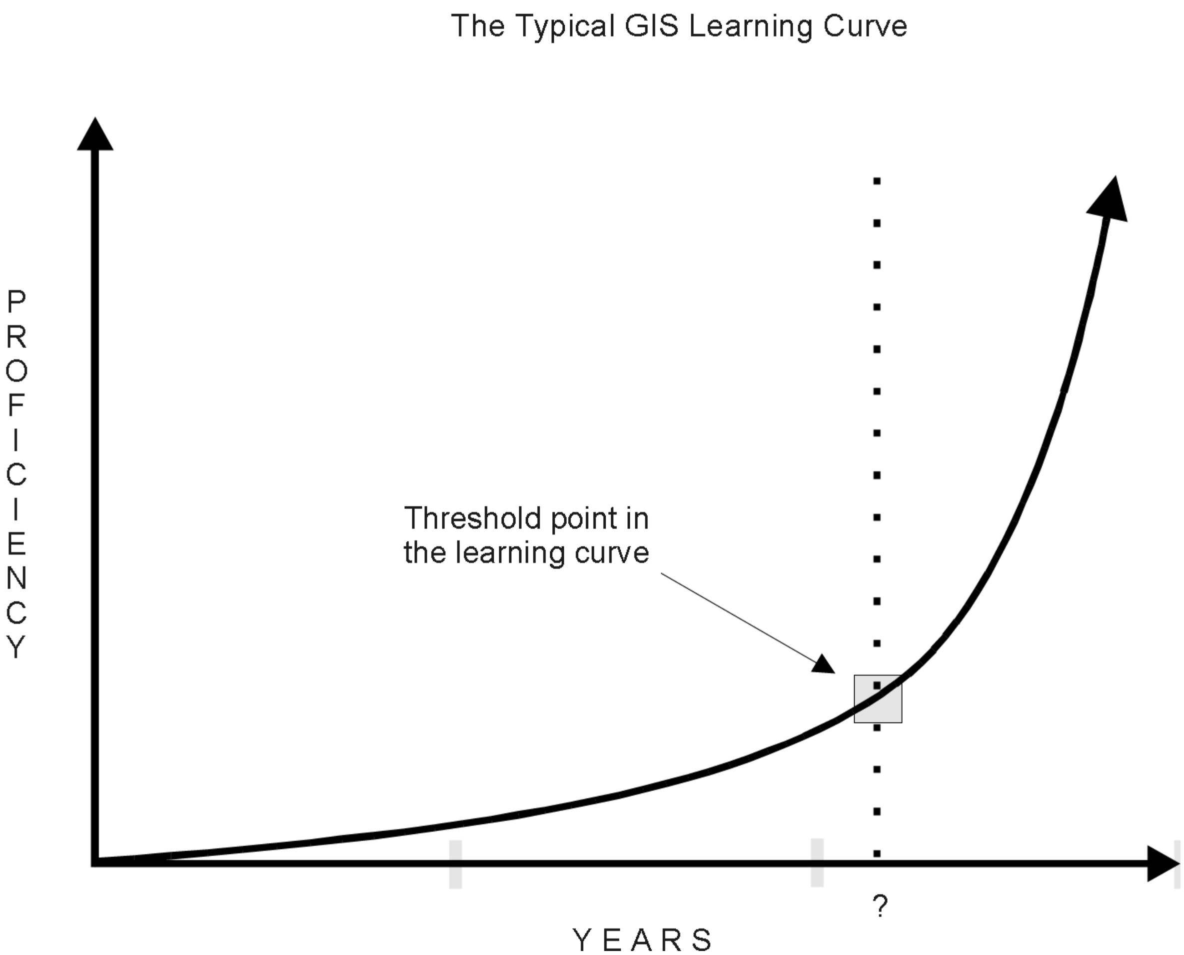
  Training and support for on-going learning, for in-house staff as well as new personnel, is essential for a successful implementation. Staff at the three levels should be educated with respect to the role of the GIS in the organization. Education and knowledge of the GIS can only be obtained through on-going learning exercises. Nothing can replace the investment of hands on time with a GIS !

8.4.2 The Learning Curve

Contrary to information provided by commercial vendors of GIS software, there is a substantial learning curve associated with GIS. It is normally not a technology that one becomes proficient in overnight. It requires an understanding of geographical relationships accompanied by committed hands-on time to fully apply the technology in a responsible and cost effective manner. Proficiency and productivity are only obtained through applied hands on with the system ! GIS is an applied science. Nothing can replace the investment of *hands-on* with GIS. The following figure presents the typical learning curve for GIS installations.

The learning curve is dependent on a variety of factors including :

- the amount of time spent by the individual with hands-on access;
- the skills, aptitude and motivation of the individual;
- the commitment and priority attached to GIS technology dictated by the organization and management;
- the availability of data; and
- the choice of software and hardware platforms.



A critical requirement for all GIS implementations is that adequate education and training is provided for operational staff, as well as realistic priorities are defined with which to learn and apply the technology. This is where a formal training curriculum is required to ensure that time is dedicated to learning the technology properly. Adding GIS activities to a staff member's responsibilities without establishing well defined milestones and providing adequate time and training mechanisms is prone to failure. A focused and properly trained operations staff that has consistent training will result in greatly reduced turnaround times for operations, and ensure consistency in quality of product.

The threshold point of the learning curve is typically around the two year time frame. However, this is dependent on the ability of the organization to establish a well defined and structured implementation plan that affords appropriate training and resources for technical staff. The flat part of the learning curve can be shortened if proper training is provided, data is available for use, the *right* software and hardware is acquired.

The typical learning curve reflects a long initial period for understanding spatial data compilation requirements and database architecture. However, after data models are well understood and sufficient data compilation has been completed the learning curve accelerates. Once a formal application development environment is established and user needs are well defined an infrastructure exists for effective application of the technology. Building operational applications based on formal functional specifications will result in continued accelerated learning. The *data hurdle* is often a stumbling block for many GIS users.

8.4.3 The Productivity Curve

GIS is a long term investment that matures over time. The turnaround for results may be longer than initially expected. The establishment of a formal implementation strategy will help to ensure that realistic expectations are met. Data is the framework for successful application of GIS technology. In this respect, the investment in establishing a solid data platform will reap rewards in a short term timeframe for establishing a cost-effective and productive GIS operation. The availability of quality data supplemented by a planned implementation strategy are the cornerstones of achieving a productive and successful GIS operation. A robust database should be considered an asset !

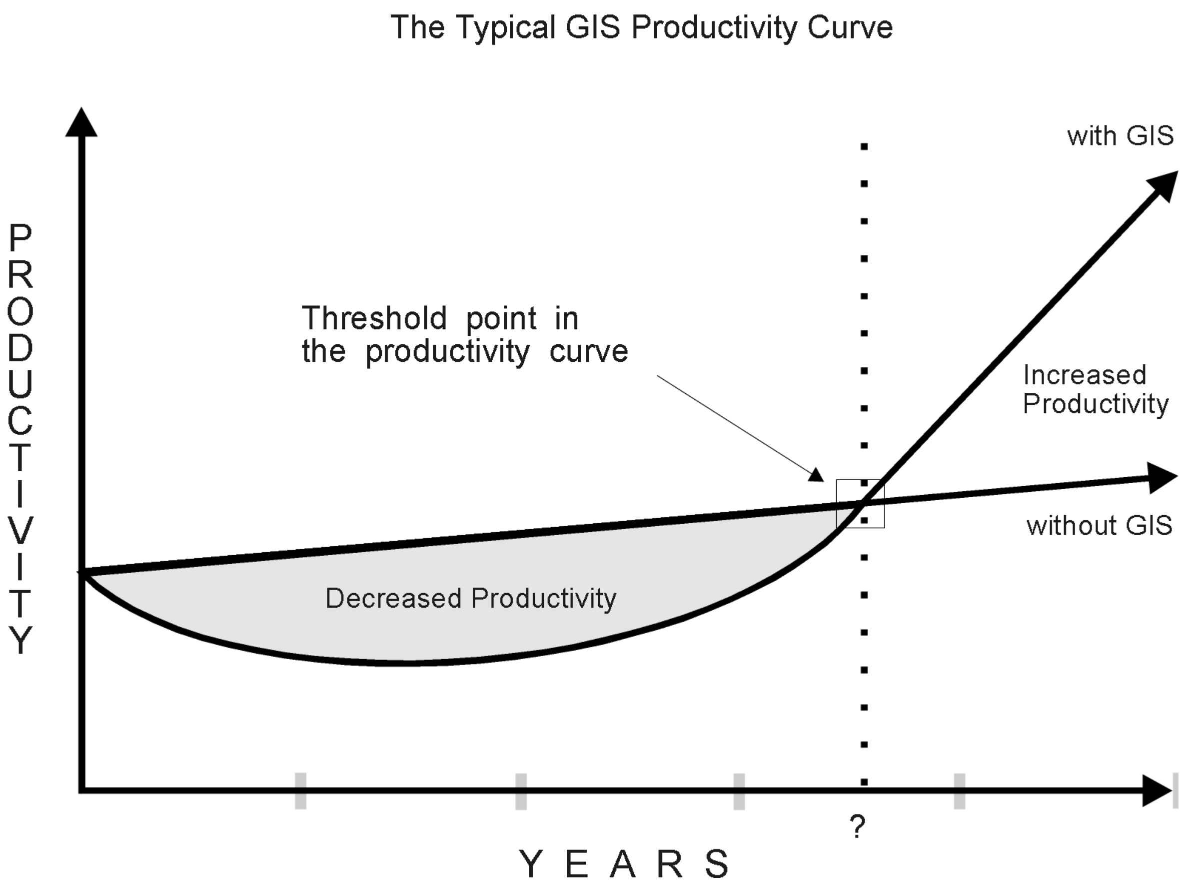
However, even with a well defined and systematic implementation strategy GIS technology will not provide immediate benefits. Benefits and increased productivity are not achieved overnight. GIS technology is complex in nature, has a generally steep learning curve, and requires a *complement of skills* for it to be applied successfully. In fact, most organizations realize a loss in overall operational productivity over the short term while the GIS platforms are being installed, staff is trained, the learning curve is initiated, and data is being captured. This is common of all office automation activities. The following figure presents the typical situation that occurs with respect to comparing long term productivity with, and without, GIS technology.

Depending on the unique circumstances of the implementation process, the status of data compilation, and the organizational climate, increased productivity is normally reached between the second and fifth year of implementation. This is identified by the threshold point. Again, this is dependent on a variety of factors including :

- the skills and experience of the staff involved;
- the priority and commitment by the organization;
- the implementation strategy; and
- the status of data compilation.

The primary issue with implementing GIS is to achieve the threshold point of increased productivity in the shortest possible time frame. In other words, minimize the time in which a decrease in productivity occurs. Of course, the issue of productivity is typically of greaterst concern with private industry, e.g. forestry companies. Nonetheless, the significant investment in hardware/software, data, and training necessitates that a structured approach be utilized to achieve the threshold point in the shortest possible time frame.

A GIS acquisition based on well defined user needs and priorities is more likely to succeed than without. A major pitfall of most installations with GIS technology, e.g. particularly forestry companies and government agencies, is the lack of well defined user needs on which to base the GIS acquisition and implementation.



8.4.4 The Implementation Plan

Implementation can be seen as a six phase process.[46] They are :

- **Creating an awareness**

GIS needs to be *sold* within an organization. The education of staff is very important. Depending on the way in which GIS technology is being introduced to the organization the process for creating an awareness may differ. Technical workshops are often appropriate when a *top-down* approach exists, while management workshops are often more relevant when a *bottoms-up* approach exists. Education of the new technology should focus on identifying existing problems within an organization. These often help justify a GIS acquisition.

[46] Aronoff provides an excellent review of general implementation phases.

They include :

- spatial information is poorly maintained or out of date;
- spatial data is not recorded or stored in a standard way;
- spatial data may not be defined in a consistent manner, e.g. different classifications for timber information;
- data is not shared between departments within an organization;
- data retrieval and manipulation capabilities are inadequate to meet existing needs; and
- new demands are made on the organization that cannot be met with existing information systems.

Identifying System Requirements

The definition of system requirements is usually done in a *user needs analysis*. A user needs analysis identifies users of a system and all information products required by those users. Often a prioritization of the information products and the data requirements of those products is also undertaken. A proper user needs analysis is crucial to the successful evaluation of GIS software alternatives.

After user needs have been identified and prioritized they must be translated into *functional requirements*. Ideally, the functional requirements definition will result in a set of processing functions, system capabilities, and hardware requirements, e.g. data storage, performance. Experienced GIS consultants often play a major role in this phase.

System Evaluations

Evaluating alternative hardware and software solutions is normally conducted in several stages. Initially a number of candidate systems are identified. Information to support this process is acquired through demonstrations, vendor literature, etc. A short listing of candidates normally occurs based on a low level assessment. This followed by a high level assessment based on the functional requirements identified in the previous phase. This often results in a rating matrix or template. The assessment should take into account production priorities and their appropriate functional translation. After systems have been evaluated based on functional requirements a short list is prepared for those vendors deemed suitable. A standard benchmark, as discussed earlier, is then used to determine the system of choice.

Justifying the System Acquisition

The proper justification of the chosen system requires consideration of several factors. Typically a *cost-benefit analysis* is undertaken to analyze the expected costs and benefits of acquiring a system. To proceed further with acquisition the GIS should provide considerable benefits over expected costs. It is important that the identification of intangible benefits also be considered.

The justification process should also include an evaluation of other requirements. These include data base development requirements, e.g. existing data versus new data needs and associated costs; technological needs, e.g. maintenance, training, and organizational requirements, e.g. new staff, reclassification of existing job descriptions for those staff who will use the GIS.

- **System Acquisition and Start Up**

After the system, e.g. hardware, software, and data, is acquired the start up phase begins. This phase should include *pilot projects*. Pilot projects are a valuable means of assessing progress and identifying problems early, before significant resources have been wasted.[47] Also, because of the costs associated with implementing a GIS it is often appropriate to generate some results quickly to appease management. First impressions are often long remembered.

- **Operational Phase**

The operational phase of a GIS implementation involves the on-going maintenance, application, and development of the GIS. The issue of responsibility for the system and liability is critical. It is important that appropriate security and transaction control mechanisms be in place to support the system. A systematic approach to system management, e.g. hardware, software, and data, is essential.

8.5 Future Developments and GIS Trends

The development and application of geographic information systems is vibrant and exciting. The term GIS remains one of the most popular *buzz words* in the computer industry today. GIS is perceived as one of the emerging technologies in the computer marketplace. The involvement of major computer vendors is an illustration of this fact. Everybody wants a GIS. This popularity is not without its validity however. GIS is very much a multi-disciplinary tool for the management of spatial data. It is inherently complex because of the need to integrate data from a variety of sources. Functions must accommodate several application areas in a detailed and efficient manner. A variety of important developments are occurring which will have profound effects on the use of GIS. They are identified in the following sections.

8.6.1 New Data Sources

The generation of data from new sources is an on going development. Application specialists have traditionally attempted to research and implement new data sources into their work. Most of these new data sources are based strictly on scientific technological developments.

Remote sensing is will become, if it is not already, the primary source for new data. Due to recent technological developments in hardware most GIS software can now accommodate remotely sensed imagery at high resolutions, and in varying formats. Remote sensing data can include aerial photographs, satellite imagery, radar imagery, etc. Some of the past problems with using remotely sensed imagery have been the inability to integrate it with

[47] Aronoff, 1989.

other data layers, particularly vector encoded data. Remote sensing specialists stress that their data is of most value when combined with, and substantiated by, other data sources. Several commercial GIS products are now offering their software *bundled* with an *image processing* software package. Many of these packages allow you to interactively view data from both systems simultaneously, and also afford the conversion of data between systems. The integration of GIS and image processing capabilities offers a great potential for resource specialists.

Another data source that has generated much interest is **Digital Elevation Models (DEM)**. Elevation data has traditionally been generated from the interpolation of contour information. However, recent technological developments and the establishment of several digital mapping projects by government agencies has propagated the use of and interest in elevation modelling. Several different sources of DEM data exist within Canada. The most common and readily available DEM data can be acquired from either the federal government, e.g. 1:250,000 map scale, or from selected provincial government agencies. For example, DEM data commensurate with a 1:20,000 map scale is distributed by the Alberta Government under the 1:20,000 Provincial Digital Mapping project. In British Columbia, DEM data is available with the 1:20,000 TRIM project. In both these cases DEM data is captured photogrammetrically during the stereo-compilation phase of the topographic data capture process. Each DEM is comprised of X,Y, and Z coordinates at regular intervals across a map sheet. This regular grid is supplemented by spot height data points and breakline information (irregular points). In the United States, DEM data is available from a variety of sources, however the most common is the USGS (United States Geological Survey) 1:24,000 QUAD sheets.

DEM data can be used in the generation of a variety of data derivatives. The most common are slope and aspect. The ability to integrate DEM data is a common function within most GIS packages. However, it is typically offered as a separate module that must be purchased individually.

8.6.2 Expert Systems

Perhaps the newest technological development that holds the greatest promise for GIS is expert systems and artificial intelligence. Research in a variety of applications have attempted to integrate artificial intelligence methodologies with GIS capabilities. The development of *rules based* approaches within selected areas of GIS functionality have been applied and successfully integrated. Feature generalization is an good example. However, most of the research in this domain is highly specialized. The highly sophisticated nature of expert systems has caused it to be somewhat removed from the mainstream of GIS applications. It is the GIS tool of the future.

8.6.3 Hardware and Software Developments

The technological advancements made in hardware and software development over the past few years have been phenomenal. The distinction between personal computer and workstation, a mainstay during the 1980's has become very fuzzy. Recent developments within the micro-chip industry, e.g. the Pentium chip, have made the micro-computer a viable and promising tool for the processing of spatial data. Most notable of these is the emergence of 32-bit Pentium chip micro-computers and the use of the Windows NT operating environment.

Several trends in hardware and software development for GIS technology stand out. These are reviewed below :

- The replacement of dumb terminals with intelligent workstations;
- The dominant hardware system architecture for GIS systems during the 1980's was the centralized multi-user host network. The distributed network architecture, utilizing UNIX based servers, and desktop workstations, has been the norm over the past five years.;
- The trend in disk storage is towards greatly increased storage sizes for micro-computers, e.g. PC's and workstations, at a lower cost;
- The emergence of relatively low cost reliable raster output devices, in particular inexpensive ink jet based plotters, has replaced the more expensive color electrostatic as the ad hoc standard plotting device for GIS.;
- The emergence of fast, relatively inexpensive micro-computers with competitive CPU power, e.g. 32-bit Penitum has challenged the traditional UNIX stronghold of GIS.;
- While the de facto operating system standard has been UNIX for several years with GIS, the Windows NT operating system is emerging as a serious and robust alternative. This is especially prevalent with organizations wishing to integrate their office computing environment with their GIS environment. This trend is closely associated with the development of 32-bit micro-computers.;
- SQL (Standard Query Language) has become the standard interface for all relational DBMS;
- The ability to customize user interfaces and functionality through Application Programming Interfaces (API) and macro languages. The major development in GIS technology over the past five years has been the ability to customize the GIS for specific needs. Application development is a mandatory requirement for all GIS sites, and should be weighted accordingly when considering a GIS acquisition.

APPENDIX A - Glossary of Terms

Accuracy Conformance to a recognizable standard. Correctness. The resolution rating that determines the quality or fitness for use of selected data. Typically referred to in geographic data management by map scape, e.g. 1:20,000, 1:100,000.

AFS Alberta Forest Service.

AISD Automated Information Systems Division.

ALGORITHM

A set of rules for solving a problem, most often involving a set of mathematical calculations.

API Application Programming Interface. A set of programming tools provided with commercially available softwere that afford customization of the software to satisfy explicit user requirements.

APPLICATION

A particular type of project undertaken by a resource specialist.

ARC A line connecting a set of points terminating in a node. Often forming one side of a polygon.

AREA A fundamental unit of geographic information. The extent or measurement of a surface (polygon).

ASPECT Azimuth of the maximum slope. Used in DEM applications. Aspect is usually given in 8 directions N,NE,E,SE,S,SW,W,NW.

Attribute Non-graphic information associated with a point, line, or polygon feature in a GIS. Typically a characteristic of the feature stored in a relational database table.

ATS Alberta Township System. The basic survey system for the province of Alberta.

AVI Alberta Vegetation Inventory. The prototype vegetation inventory of 500 townships that were compiled in digital form for selected areas of Alberta. The precursor to CVI and VIS.

Benchmark Test

A test to evaluate the capabilities of a computer system in terms of the customer's requirements and performance.

BOOLEAN Operations

Logical operations performed on sets. The result of Boolean operations is either TRUE or FALSE.

Buffering

Commonly used in vector operations to create a distance band, or buffer, around a selected feature.

CAD Computer Aided Drafting. The graphics or mapping segment of the geographic information management process. CAD is the technology utilized in computer-assisted mapping.

Configuration

A particular combination of computer hardware and software for a certain class of application tasks.

Contour A line connecting points of equal elevation (isoline).

Coverage Referring to the graphic and attribute information related to a particular study area. Also, the term utilized by the ARC/INFO software to identify a specific data layer.

CVI Collective Vegetation Inventory. AVI with revised specifications and formats to accomodate private industry and government users.

Database A collection of interrelated information, usually stored on some form of mass-storage system. A GIS data base includes data about the position and attributes of geographic features.

DBMS Data Base Management System. The subsystem used to store attribute data in a GIS.

Device A piece of equipment external to the computer designed for a specific function such as data input, data storage, or data output.

Digital The ability to represent data in discrete, quantized units as is typical to computer formats.

DEM A quantitative model of topography in digital form. Sometimes called digital terrain model (DTM). A DEM contains elevation values, usually a regular grid of points, that describe the surface of a study area. Typically DEM points contain an X and Y coordinate to located the point on the earth, and a Z coordinate in metres above sea level.

Digitizer A device for entering the spatial coordinates for mapped features from a map or document in a computer.

Dissolve

A function commonly used is GIS to remove boundaries between areas of common ratings after reclassification.

Distributed Processing

The placement of hardware processors where needed, e.g. on nodes of a network, instead of concentrating all computing power in a large central CPU.

Drape Overlay in a draping fashion graphic features onto a DEM surface.

Elevation angle

The angle above the surface from which a perspective view will be generated.

ERD Entity-Relationship Diagram. A method used to model business, organization, or government data requirements.

ESRI Environmental Systems Research Institute. Developers of ARC/INFO GIS software package.

Format The way in which data are systematically arranged for transmission between computers.

FMA Forest Management Agreement

Geo-administrative Boundaries

Geographic based administrative boundaries typically used in natural resourve management, e.g. Forest Management Unit, Wildlife Management Unit, etc.

GIS Geographic Information System

GKS Graphics Kernel System. A set of software primitives for allowing device-independent graphics programming.

Hardcopy A paper copy of a map or document

Hardware The physical components of a GIS - the computer, plotter, printers, VDT's.

Information Product

An output product, graphic, map or report, that is generated by a module of the system in support of decision making processes.

Input The data entered into a computer system.

Interactive A computer action or tasks which is initiated by the operator to in real time.

Interface A hardware-software link that allows two computer systems to be connected for data communication. (See also, User Interface)

Interpolation

The process of determining the value of an unsampled data point for a given X,Y location based on the values of surrounding sampled data points.

IOS Integrated Office Systems

Labels Commonly used as the unique identifier for geographic features, e.g. a label point for a polygon. Distinguished from annotation (text) as it is unique for every feature.

Layer A logical separation of mapped information in a vertical manner according to themes, e.g. vegetation, soils, transportation, survey fabric, hydrography, etc.

Line A basic geographic unit, defined by at least two vertices.

LISD Land Information Systems Division. Alberta Forestry, Lands and Wildlife.

Map Projection

A mathematical transformation that is used to represent a spherical surface on a flat map.

Modelling (Spatial)

The studying of landform processes using quantitative / mathematical algorithms and/or models.

Node The point at which arcs in a polygon or linear network are joined. Also the distinct point in a feature where lines end. Required in vector GIS to define topology.

NRIS Natural Resources Information System

Offering A set of computer programs that can be used for a particular generalized class of applications. Often referred to as a software package.

Output The results of processing data in a GIS.

Package A set of computer programs than can be used for a particular generalized class of applications.

PFS Proposal For Solutions phase of a structured system development methodology.

Phase 3 Inventory

Phase 3 forest inventory circa 1980. The most recent complete forest inventory for the province of Alberta. Phase 3 is a manual inventory composed of typical forest inventory 1:15,000 map sheets.

Polygon A closed, areal boundary representing an area of homogeneous characteristics on a map. Typically, the method to encode thematic data in a GIS.

Raster Refers to an area partitioned into a set of grid cells (usually square) with each cell assigned a value describing some spatially related characteristic. Typically a raster structure is defined in rows and columns. GIS's utilize vector and raster data structures for different functionality.

Resolution The smallest spacing between two display elements. Resolution can be spatial, e.g. 10 metres, or defined by pixel size, e.g. 1024 x 768 pixels.

RDBMS Relational Data Base Management System. A data base offering whose architecture and functionality adhere to the rules as defined by the relational data base model. RDBMS is the accepted norm for database in the GIS area.

RIB Resource Information Branch, Land Information Services Division.

Scale The relation between the size of an object on a map and its size in the real world. Scale is the basic denominator for defining map accuracy.

Sliver A gap between two lines created erroneously during processing. Commonly small polygons formed during building of topology.

Slope Gradient of elevation. The most common derivative in DEM modelling.

Spaghetti digitizing

Refers to the digitizing of map features without any regard to the sequence or identification of line/point intersections.

Spatial Indexing

The partitioning and indexing of a spatial study area to optimize feature search and retrieval operations in a horizontal fashion. Spatial indexing is a necessary internal requirement for a GIS to handle large data sets that cover a large study area.

SQL Structured Query Language. An ANSI (American National Standards Institute) standard language used to query relational data bases.

Thematic Layer

Data associated with one specific theme, e.g. vegetation, soils, slope, ownership, etc. See LAYER.

Thematic Products

Information products that conform to thematic classifications, e.g. vegetation classified map such as tree height or age.

Topology The mathematical definition of feature storage utilized in vector GIS. Topology has specific connectivity requirements for point, linear, and polygonal features.

UNIX An operating system utilized for workstation technology. UNIX is the accepted standard for workstation and scientific computing. UNIX includes several different *flavours* for different vendors, e.g. DEC-ULTRIX, IBM-AIX, Intergraph-CLIX, etc.

Vector A means of coding geographic features by storing explicit coordinates. A type of GIS data structure for spatial data.

Vertical Exaggeration

A constant value used to multiply Z values in a DEM for realistic perspective viewing.

Viewing Azimuth

The viewing location for a 3-d perspective view. Normally measured in degrees from a 0 degree axis.

VIS Vegetation Information System. The system to be used to store existing AVI and new CVI data. VIS will also support the generation of basic information products from this data.

Workstation Microcomputer based on either a Complex Instruction Set Chip (CISC), or a Reduced Instruction Set Chip (RISC). The fundamental hardware currently utilized for geographic information management and GIS development.

APPENDIX B - References and Information Sources

The following references were used extensively in the preparation of this book and are suggested reading for those who wish to obtain more information about GIS technology.

Anderson, A.; Starr, L. "*Geographic Information Systems*", in South African Journal of Photogrammetry, Remote Sensing, and Cartography, Vol. 14, June 1984.

Aronoff, Stan. *Geographic Information Systems : A Management Perspective*, Ottawa: WDL Publications, 1989.

Aronson, Peter. "*Attribute Handling for Geographic Information Systems*", in Proceedings of Auto-Carto 8. Baltimore, Maryland, April 1987.

Berry, Joseph K. *Beyond Mapping: Concepts, Algorithms, and Issues in GIS*. GIS World Publications, 1993.

Berry, Joseph K. *Spatial Reasoning for Effective GIS*. GIS World Publications, 1995.

Buckley, D.J. *Standard Benchmark for the Evaluation of Micro-computer GIS Software*. Prepared for the Alberta Forest Service, February 1986.

Buckley, D.J.; Pawlina, M.W.; Strickland, R. *Automation of Visible Area Mapping for Fire Detection Lookouts*. A paper presented at the GIS'90 Symposium, March 1990, Vancouver, B.C.

Burrough, P.A. *Principles of Geographical Information Systems for Land Resource Assessment*, New York: Clarendon Press, 1986.

Carter, James. "*Digital Representations of Topographic Surfaces*", in PE & RS, Nov. 1988, P.1577-1580.

Chrisman, Nicholas. "*Fundamental Principles of Geographic Information Systems*", in Proceedings of Auto-Carto 8. Baltimore, Maryland, April 1987.

Cooney, T. "*Gee Whiz ... GIS !*", in The Compiler, a publication of the Forest Resources Systems Institute, Florence, Alabama, Fall 1986. P.9-34.

Cowan, David J. "*GIS versus CAD versus DBMS : What Are the Differences ?*", in PE & RS, Vol. 54, No. 11, Nov. 1988, P. 1551-1555.

Croswell, Peter; Clark, Stephen. "*Trends in Automated Mapping and Geographic Information System Hardware*", in PE & RS, Nov. 1988, P.1571-1576.

Dangermond, Jack. "*GIS Trends and Experiences"*, Keynote Address, Proceedings of the 2nd International Symposium on Spatial Data Handling, Seattle, Wash., 1986.

Dangermond, Jack. "*Trends in Hardware for Geographic Information Systems*", in Proceedings of Auto-Carto 8. Baltimore, Maryland, April 1987.

Dangermond, Jack. "*A Review of Digital Data Commonly Available and Some of the Practical Problems of Entering Them in a GIS*". Source unknown. 1987.

Dangermond, Jack. "*GIS Trends and Comments*", An Editorial in the ARC/INFO News, Summer/Fall 1988. P.13 - 17.

Dillencourt, M.B.; Samet, H. "*Extracting Region Boundaries from Maps Stored as Linear Quadtrees*", in Proceedings of the Third International Symposium on Spatial Data Handling, Sidney, Australia, August 1988.

Frank, Andrew. "*Requirements for a Database Management System for a GIS*", in PE & RS, Nov. 1988, P.15557-1564.

Goodenough, David. "*Thematic mapper and SPOT Integration With a Geographic Information System*", in Journal of Photogrammetric Engineering and Remote Sensing, Vol. 54, No.2, February 1988, P.167-176.

Hunter, G.J. Williamson, I.P. "*The Need for A Better Understanding of the Accuracy of Spatial Databases*", in Proceedings of the 1990 URISA Conference, Edmonton, Alberta. October 1990.

Jensen, S.K.; Dominque, J.O. "*Extracting Topographic Structure from Digital Elevation Data for GIS Analysis*", in PE & RS, Nov. 1988, P.1593-1600.

Jordan, G. "*Making GIS Work in Forest Management*", in Proceedings of GIS'89, an International Symposium on Geographic Information Systems. Vancouver, B.C. March 1989.

Lupien, A.; Moreland, W.; Dangermond,J. "*Network Analysis in Geographic Information Systems*", in Photogrammetric Engineering and Remote Sensing, Vol. 53, No. 10, October 1987, P. 1417-1421.

Marble, D.; Peuquet, D. "*Geographic Information Systems and Remote Sensing*", in Manual of Remote Sensing, 2nd Edition, Falls Church, Virginia. ASP, P.923-958.

Marble, Duane. "*GIS and LRIS: Differences and Similarities"*, in FIG: International Symposium on Land Information Systems Proceedings, Edmonton, Alberta. October 1984.

Muller, J.C. *Spatial Accuracy of Digital Thematic Maps in the Context of the 1/20000 Soil Inventory Project*. A report prepared under contract to the Resource Information Branch, Alberta Forestry, Lands & Wildlife. Edmonton, Alberta. March 1987.

Newcomer, J.A., Szarjgin, J. "*Accumulation of Thematic Map Error in Digital Overlay Analysis*", in The American Cartographer, 1984, P.58-62.

Noronha, V.T. "*A Survey of Hierarchial Partitioning Methods for Vector Images*", in Proceedings of the Third International Symposium on Spatial Data Handling. Sidney, Australia. August 1988.

Olson, Judy. "*Comments on Research Trends"*, in The American Cartographer, A Special Issue: U.S. National Report to ICA, 1984, J. Olson ed., Summer 1984.

Parker, H. Dennison, "*The Unique Qualities of a Geographic Information System: A Commentary*", in Photogrammetric Engineering
and Remote Sensing, Vol.54, No. 11, P.1547-1549, November 1988.

Pawlina, M., Buckley, D., Strickland, R., "*The Automation of Visible Area Mapping for Fire Detection Lookouts*", A paper presented at the GIS'90 Symposium, March 1990.

Peuquet, Donna, "*A Conceptual Framework and Comparison of Spatial Data Models*", in Cartographica, Vol. 21, No. 4, 1984, P.66-113.

Robinson, A.; Sale, R.; Morrison, J. *Elements of Cartography*, 4th Edition. New York: John Wiley and Sons, 1978.

Stiefel, Malcolm. "*Mapping Out the Differences Among Geographic Information Systems*", in The S.Klein Computer Graphics Review, Fall 1987, P.73-87.

Tomlinson, R.; Calkins, H.; Marble, D. C*omputer Handling of Geographic Data: An Examination of Selected Geographic Information Systems*, Paris: Natural Resources Research Series XIII, The UNESCO Press, 1976.

Vitek, J.D.; Walsh, S.J.; Butler, D.R. "*Accuracy in Geographic Information Systems: An Assessment of Inherent and Operational Errors*", in Proceedings PECORA IX Symposium, 1984. P.296-302.

Walsh, S.J.; Lightfoot, D.R.; Butler,D.R., "*Recognition and Assessment of Error in Geographic Information Systems*", in Journal of Photogrammetric Engineering and Remote Sensing, Vol. 53, No. 10, October 1987, P.1423-1430.

Wilson, P. "*Multiple Purpose GIS*", in URISA : Papers of the Annual Conference, 1984, P.137-143.

There are several journals that regularly have relevant GIS articles etc. They are listed below :

International Journal of Geographic Information Systems, Published quarterly by Taylor & Francis Ltd.

Cartography and Geographic Information Systems, The Journal of the American Congress on Surveying and Mapping. Published quarterly.

Photogrammetric Engineering & Remote Sensing, Published monthly by the American Society for Photogrammetry and Remote Sensing.

Cartographica, The Journal of the Canadian Cartographic Association. Published quarterly.

Several magazines dedicated to providing information on the GIS marketplace and applications exist. They include :

GIS World. News of Geographic Information System Technology in Land, Natural Resources, & Urban Information Management. GIS World Inc., P.O. Box 8090, Fort Collins, Colorado, USA 80526.

Earth Observation Magazine. EOM Inc. P.O. Box 3623, Littleton, CO 80161. Phone (303) 909-6619.

Geo Info Systems. **Applications of GIS and Related Spatial Information Technologies.** Aster Publishing Corp. 859 Williamette St. P.O. Box 10460, Eugene, Oregon 97440-2460. Phone (503) 343-1200.

The Compiler. A Forest Resources Systems Institute (FORS) Publication. 122 Helton Court, Forence, AL 35630. Telephone: (205) 767-0250.