IAEA-TECDOC-782

# *Spatial data integration for mineral exploration, resource assessment and environmental studies: A guidebook*

INTERNATIONAL ATOMIC ENERGY AGENCY IAEA

December 1994

The IAEA does not normally maintain stocks of reports in this series. However, microfiche copies of these reports can be obtained from

INIS Clearinghouse
International Atomic Energy Agency
Wagramerstrasse 5
P.O. Box 100
A-1400 Vienna, Austria

Orders should be accompanied by prepayment of Austrian Schillings 100,– in the form of a cheque or in the form of IAEA microfiche service coupons which may be ordered separately from the INIS Clearinghouse.

The originating Section of this document in the IAEA was:

Nuclear Materials and Fuel Cycle Technology Section
International Atomic Energy Agency
Wagramerstrasse 5
P.O. Box 100
A-1400 Vienna, Austria

SPATIAL DATA INTEGRATION FOR MINERAL EXPLORATION,
RESOURCE ASSESSMENT AND ENVIRONMENTAL STUDIES: A GUIDEBOOK
IAEA, VIENNA, 1994
IAEA-TECDOC-782
ISSN 1011–4289



Printed by the IAEA in Austria
December 1994

## FOREWORD

The International Atomic Energy Agency has played a significant role over the years in the improvement and use of uranium exploration techniques and data obtained through uranium exploration. Numerous documents on uranium geology and exploration methods have been published.

The purpose of this document is to provide an introduction to the new tools and applications of computer based spatial data integration as used by geologists. During mineral exploration vast amounts of spatial data are collected. Some of these data are gathered routinely from satellite and airborne sensors in a digital form. Large sets of ground survey data are also stored in digital form. Integrated use of these data has in the past been carried out qualitatively and visually. Recent advances in personal computing give the possibility to integrate several layers of quantitative data and produce the images that are useful in the overall interpretation of the combined information, not only in the uranium exploration field, but also in environmental studies.

In order to provide the experts involved in uranium exploration with information on recent developments in computer applications for spatial data integration and image processing, the IAEA convened consultants meetings in November 1991 and November 1992 to produce this guidebook, which contains information on spatially distributed data, data capture, database creation and visualisation of data. Vector, and in particular, raster data types and the aspects of integration modelling are discussed. For better understanding of the subject several case studies are included in the document. Guidance on the specification of technical requirements and the staffing needed to undertake spatial data processing projects is also provided.

The IAEA wishes to thank the consultants who took part in the preparation of the guidebook. They included G.F. Bonham-Carter of the Geological Survey of Canada, R.B. McCammon of the United States Geological Survey, A.G. Fabbri of the International Institute for Aerospace Survey and Earth Sciences, Netherlands, and V. Kuosmanen, of the Geological Survey of Finland. Credit should also be given to E.M. Schetselaar and M.A. Groossens, Netherlands, who contributed to the preparation of the case studies. The IAEA staff member responsible for the project was M. Pecnik of the Division of Nuclear Fuel Cycle and Waste Management.

## *EDITORIAL NOTE*

*In preparing this document for press, staff of the IAEA have made up the pages from the original manuscript(s). The views expressed do not necessarily reflect those of the governments of the nominating Member States or of the nominating organizations.*

*Throughout the text names of Member States are retained as they were when the text was compiled.*

*The use of particular designations of countries or territories does not imply any judgement by the publisher, the IAEA, as to the legal status of such countries or territories, of their authorities and institutions or of the delimitation of their boundaries.*

*The mention of names of specific companies or products (whether or not indicated as registered) does not imply any intention to infringe proprietary rights, nor should it be construed as an endorsement or recommendation on the part of the IAEA.*

## CONTENTS

# 1. INTRODUCTION

## 1.1. PURPOSE OF DOCUMENT

The past decade has seen radical changes to man's ability to collect, store and analyze spatial data with computers. Digital images are routinely gathered from a host of space- and airborne sensors; large suites of samples of geological media are chemically analyzed and the data stored digitally, often for thirty or more elements simultaneously; even geological mapping can now be computerized in the field. Advances in computer technology have not only spurred the collection of huge volumes of spatially-referenced data, but also provided the means to store, manipulate, visualize and analyze these data. Fast, relatively inexpensive personal computers equipped with image processing (IP) and geographical information system (GIS) software have the potential to bring about radical changes in the geological workplace. Instead of overlaying paper maps by hand on a light table, maps and images can be electronically combined. This is not only far more efficient for examining spatial associations between spatial data layers, but it also allows more exhaustive and creative use to be made of expensively-collected data.

Two areas of application where this new technology are particularly important are mineral resources and environmental studies. Although the goals of these fields differ, one being the discovery and assessment of mineral deposits, the other being the understanding and impact of man-made changes to the earth, both require the simultaneous assessment and combination of multiple layers of geological, geophysical, geochemical and associated data. The purpose of this document is to provide an introduction to the tools and applications of computer-based ***spatial data integration***, as used by geologists.

Until recently, most scientists and engineers would use the services of a typist to prepare a report or manuscript, yet now many find that they can do the job themselves with a word-processing package on a personal computer. In addition, spreadsheets, statistical packages, drawing programs and others are routinely used for analyzing and presenting data. This has come about not only because of advances in computer hardware, but also because software has become much friendlier and easy to use. At the same time, however, the person who can apply these methods effectively must acquire new technical skills, and such training takes time and effort. This is particularly true of computer systems for manipulating and integrating spatial data. It may take only an hour or two to learn how to use a 'desk-top mapping' program to manipulate and display maps and statistics with a prepared database. But much more training is required to use a GIS to build a spatial database from scratch, and to use it creatively to carry out a mineral exploration study or an environmental assessment. Furthermore, papers on modern methods of spatial data integration are scattered in the literature, and there is a need for a document to bring a number of diverse topics together and present them in a form that will serve as a guidebook to the subject.

This document is a general overview of the subject for the non-specialist, and is also an introduction to some technical aspects for those who wish to apply the methodology themselves. As often happens in new technology fields, the training tends to be centred around a particular computer product, and the student is required to wrestle with thick manuals that deal with very specific tasks at the expense of viewing the problem as a whole. Here we have avoided as much as possible the jargon and myopia of technical manuals, and have tried to provide a general guidebook, as suggested in the title. The document could be used effectively in a GIS course, in conjunction with laboratory exercises using a specific GIS and/or image processing software.

## 1.2. ROLE OF GEOGRAPHICAL INFORMATION SYSTEMS

One of the most significant developments in the computer handling of spatial data is the rise of what are now known as geographical information systems (GISs). There are many definitions of GISs, but essentially a GIS is a computer system (both hardware and software) for capture, storage, manipulation, visualization and analysis of geographically-referenced data. GISs have been widely applied to a host of disciplines such as resource assessment, municipal planning, transportation, marketing, forestry, epidemiology, and many others. The common factor is simply the use of information about phenomena that are distributed over the earth. Because the market for GISs is enormous and growing, some very sophisticated software has been, and is continuing to be, developed that is available at modest cost. This has led to the application of GISs in a variety of geological applications, including mineral exploration and mineral resource assessment (Knox-Robinson, 1992). GISs have not yet, however, been widely exploited by the geological community (Wadge and Pearson, 1991). One of the reasons for this is that most GISs are confined to handling 2-D data; some 3-D GISs have been developed ( Raper, 1989; Turner, 1992; Pflug and Harbaugh, 1992), but they are not yet widely used, except by some oil companies for analysis of sedimentary basins, and by some mining companies for mine planning and development. Moreover, many of the problems faced by geologists using regional data are confined to two dimensions, and 2-D maps have been the traditional medium of information storage and communication.

Most geologists are more familiar with image processing (IP) systems than with GISs, because IP systems have been around somewhat longer, and they have been widely used for display and analysis of satellite images. There is sometimes confusion about the relationship between GIS and IP. Image processing systems are computer systems for the input, manipulation, analysis and visualization of raster images. Images can come from a variety of sources and scales, including non-geographical subjects like medical images and images of thin sections of rocks. On the other hand, if the images are satellite images or rasterized maps of various kinds, and where the data are geocoded (i.e. the geographical location of the pixels is known), then an image processing system is a type of GIS. However, a full-fledged GIS is more than a raster image processing system, because it has additional capabilities for handling vector as well as raster data, it uses a database to store and access the attributes of spatial objects and it can transform data between a variety of data structures. An IP system is often used as a GIS where the spatial data to be manipulated consist of raster images. Many IP systems now provide functions for handling vector and tabular data, and conversely, GISs that started life as vector systems now handle raster images, so that the distinction between an IP and a GIS is blurred.

There are a wide variety of other types of computer programs for handling and manipulating spatial data, many of them available for personal computers. We will briefly mention computer-aided drawing or design (CAD) systems, desktop mapping systems and contouring packages. Although these programs can be very useful in their own right, any one package on its own does not provide the same range of functionality as a GIS.

CAD systems are for capturing and manipulating drawings. Point, line and polygon objects are stored in a vector format. Some very effective geological mapping software, such as FIELDLOG (Brodaric and Fyon, 1988) is based on linking CAD to a relational database management system. FIELDLOG, and other programs of its type, are suitable for field capture of geological maps and provide one type of input to a GIS. A CAD system is like a part of a vector GIS. CAD software is highly developed and has very

good display capabilities, but on its own is neither designed to carry out spatial analysis nor to use raster data types.

Desktop mapping systems are for the selective search and display of information from spatial databases. They are not designed for database-building, but can be useful for providing low-cost and effective access to an already-built GIS database.

Contouring packages are designed to take point sample data, and to produce contour maps and raster images of surfaces fitted to one or more variables. The data might be depth to bedrock at a series of well locations or gravity values, for example. There are some excellent programs of this type. This functionality is also found in GISs.

Figure 1.1 summarizes some of the various computer software packages used for spatial data handling.

## 1.3. DEVELOPING A SPATIAL DATABASE

This topic is taken up in greater detail in a separate chapter, but is introduced at this stage as a means of providing a brief overview of the typical availability and initial treatment of source data. Suppose that a mineral potential project is to be carried for a region where a geological map is available, in addition to geophysical survey data, geochemical survey data from stream sediments, and information about known mineral occurrences. Briefly let us sketch out the sequence of steps in creating a spatial database from these typical data types, showing where GIS and IP tools come into play.

***1. Geological map.*** A geological map might be digitized using a table digitizer, under the control of GIS digitizing routines, or through CAD-related software like FIELDLOG. This produces vector data, i.e. a series of point, line and polygon objects represented as strings of geographical coordinate pairs, plus information necessary for georeference purposes. Each object is also associated with various kinds of attributes. Each polygon

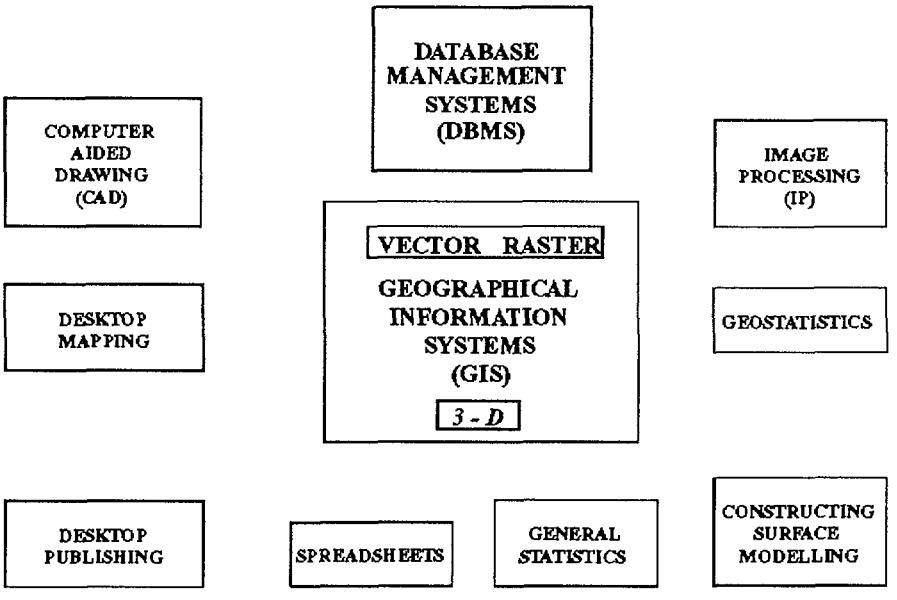



*FIG. 1.1. Types of computer software for spatial data.*

object on the geological map, for example, is described by attributes such as the name of the geological formation, its age and lithology, its metamorphic grade, type of alteration and so on. Attribute data are held in tables, and form part of a GIS database. If an IP system is being used, vector data may be converted to raster format for analysis.

*2. Geophysical surveys.* Geophysical data may originally be in the form of point measurements along profiles, such as flight-lines, or may already have been pre-processed as raster images. Specialized contouring software is used for interpolation and production of raster images, or similar routines can be found in a GIS. Each raster image is composed of a rectangular array or lattice of small elements or pixels. The geographical location of a pixel is known (or can be calculated) from its row and column coordinates. Each pixel also has a value, often a positive integer in the range 0-255, reflecting the intensity of the mapped variable. Pixel values can also act as pointers to records in an attribute table, i.e., the pixel is treated as another type of area object possessing multiple attributes. Commonly, however, a separate raster image is needed for each variable, such as total magnetic field, radiometric thorium, or Bouguer anomaly. IP systems are well suited to handling geophysical images and satellite images such as LANDSAT and SPOT.

*3. Geochemical surveys.* Geochemical data are usually available as a digital table, each record being a sample, with columns or fields containing spatial coordinates and chemical attributes. Such data can be left as point objects, linked to the table of attributes, or can be transformed into a mapped surface in various ways. The surface can be represented by interpolating on to a regular lattice, forming a raster image, with the gridding part of a contouring package. Alternatively each point can be associated with a polygon object, such as stream drainage basin, and linked to the same chemical attribute table as the sample points. This can be handled either in vector or raster mode, depending on the GIS. With an IP system, a separate raster image must be made for each geochemical element.

*4. Mineral occurrences.* Mineral deposits, showings and occurrences are usually input as a digital table, like the geochemical data, except that each occurrence remains a point object and does not form part of a piecewise or continuous surface. In vector mode, the occurrence stays as a point, linked to its attributes. In raster mode, sometimes each occurrence is converted to a pixel, forming a sparse mineral occurrence image where most of the pixels are empty and are coded with a null value.

Most of these and other data types (such as digital elevation data, drainage networks, political boundaries, land-use data, soil maps, vegetation maps) are not all available from a single source. It would be most convenient if all the data were accessible in one place ('one-stop shopping') in data formats that were consistent and well-documented. Although standards are being developed, there is no consistent set of interchange standards for spatial data in general use, at least at present. This means that assembling the available data into a consistent database, with each map layer being properly registered and in a consistent form, is a time-consuming and important step in any spatial data integration project. It would be convenient if all the relevant data was stored in one comprehensive GIS database, but building and maintaining such an information system is very expensive. At least in the near term it is unlikely that any country will assemble **all** the relevant spatial data at the scale and in the variety needed for mineral exploration into a single consistent GIS database. The cost is too great for the frequency of use.

After a specific spatial data integration project is complete, the database and the various derived maps from it are generally 'mothballed' rather than being maintained in any active status. Mothballing , or archiving, consists of transferring the whole project on to a mass storage device such as a tape, disk or possibly a compact disk (CD).

## 1.4. GENERALITY OF APPROACH

Three major steps in the spatial data integration process can be identified, see Fig. 1.2. The steps are (1) build a spatial database, (2) carry out spatial data processing and (3) apply integration models. Visualization is a particularly powerful and important activity which comes into play in both step 2 and step 3. Some of the aspects of digital data capture and data representation in step 1 have been briefly mentioned in Section 1.3. The spatial data processing of step 2 covers a wide variety of activities, dealing mainly with the pre-processing , analysis, enhancement and classification of data layers, see Chapter 3. Pre-processing includes data editing and geometric transformation to a working geographical projection, as well as transforming between data structures–vector to raster or raster to vector for example. This processing stage, coupled with the visualization of data mainly in geographical space, but also in other views such as scatterplots, variograms, power spectra and histograms, is vital for identifying spatial associations and for selecting and enhancing particular data layers to be used for the third step. Step 3 involves the application of models for combining or integrating layers. In the case of mineral exploration, the layers that are extracted and enhanced in step 2 are generally the data to be used as evidence in support of mineralization. The models for combining the evidence can be grouped into a number of types, such as data-driven versus knowledge-driven. The former are mostly statistical, whereas the latter use expert knowledge as the basis for combining spatial evidence.

One important factor that should be stressed is that a conceptual model underlies the whole spatial data integration procedure. In a sense, this conceptual model provides implicit guidelines at each step in the process, in selecting suitable data sources in step 1, in suggesting the appropriate enhancements and likely associations in step 2, and certainly for the selection and weighting of evidence in step 3. For mineral exploration and resource assessment this conceptual framework is supplied by mineral deposit models. Before discussing how deposit models are applied here, consider first the difference between mineral exploration and mineral resource assessment.

Mineral exploration is the process of locating undiscovered mineral deposits. Exploration can be carried out at various spatial scales. At a small scale, say 1:250 000, exploration using integration methods involves using all the available evidence to identify regions or tracts that have a high potential for undiscovered deposits. An exploration team uses such an approach to plan detailed fieldwork and data collection for a second stage at a larger

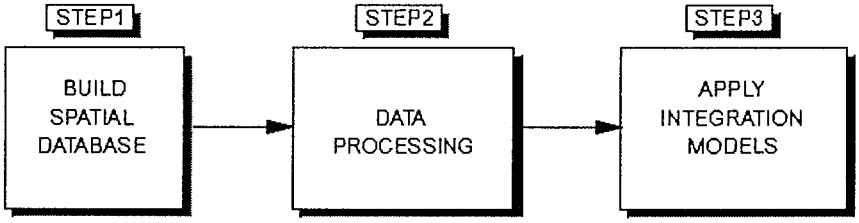



*FIG. 1.2. Spatial data integration as a three step process.*

scale, say 1:20 000. Again, regions showing elevated potential are identified, as the basis for further detailed follow-up. The final outcome, possibly after further cycles of data gathering and assessment, is the selection of drilling targets. Mineral resource assessment, on the other hand, roughly corresponds to the first stage of mineral exploration, i.e. the selection of promising regional tracts. It is also often associated with an actual prediction of the number, size and possible monetary value of mineral targets. Mineral resource assessment is carried out mostly by governments, who require the information for policy and planning. Exploration can be carried out either by the private sector or by government organizations, depending on the political and economic structure of the country concerned. The process of identifying various tracts according to their favourability for mineral deposits is also called mineral potential mapping. The application of mineral deposit models is important both in exploration and resource assessment.

Mineral deposits can be grouped together into deposit types depending on the similarity of various geological characteristics. A group of similar deposits can be represented by an ideal deposit, having all the significant characteristics that distinguish one group from another. This ideal deposit may not actually exist, but the sum total of its characteristics constitute a deposit 'model'. Many different deposit models have been identified and summarized, as illustrated by Cox and Singer (1986) and Eckstrand (1984), among others. In exploring for deposits, or assessing a region for mineralization, a number of different deposit models may be applied. In each case the diagnostic characteristics are employed both to assess the affinities of prospects of unknown type, as well as to guide the spatial integration of regional datasets. In the PROSPECTOR expert system, as described by McCammon (1989, 1990), the characteristics of deposit models are embodied in a knowledge base that is used to evaluate prospects whose attributes are elicited from a user in a question and answer session. The output is an evaluation of the most likely deposit model.

In both the traditional and computer-based spatial integration of regional datasets, deposit models play a vital role. Depending on the deposit model or models appropriate for a particular region, particular datasets are chosen for building a spatial database. The selection and enhancement of particular features, in step 2 of the integration process, is particularly affected by the deposit model under consideration. Finally, the degree of weighting of individual layers of evidence is governed by their relative importance for the particular deposit model at hand. At present, there is no body of literature comparable to that describing deposit models to guide the integration procedure for environmental studies.

## 1.5. THE ROLE OF INTEGRATION MODELLING

Having built a multi-layer spatial database, and having analyzed, selected, 'massaged', and enhanced the information into particular layers of spatial evidence to be used for prediction, or to characterize and help understand environmental phenomena, the final step is to combine the various layers together. The combination process is controlled by an integration model. An integration model is a symbolic model, using mathematical and logical symbols to combine data layers together. It usually takes the form of a set of rules, or equations, that relate a set of input maps to an output map. This kind of modelling process has also been called map, or cartographic, modelling ( Tomlin, 1990). The integration process can either take place using modelling tools within a GIS or IP system, or can be carried out in a separate customized computer program operating on the same database files.

Integration models can be a number of types. In this report we distinguish between descriptive and predictive models. Descriptive models are exploratory in nature, and combine the input maps without any specific predictive goal in mind. Predictive models, on the other hand, use combination methods that attempt to predict mineral potential, or environmental factors such as the effects of acid rain. In practice, such models are usually only partly successful at best, because of the complex nature of these problems, but they do provide consistent repeatable outcomes to a particular set of assumptions, and they are able to consider the spatial coincidence of many factors simultaneously.

It is emphasised that much of the analysis and understanding of the relationships between spatial data sources comes from their visualization. As in the field of exploratory data analysis in statistics, many of the insights gained from an integrated study come from viewing the data in different ways. This is where the graphical tools of computers are so powerful. Not only are the colour graphics important, but also the ability to explore the dataset interactively. Interactive interrogation of spatial data using the cursor and pointing to particular locations or objects, with requests about any of the data layers or their attributes, is enormously helpful. Not only can the data about a particular location be displayed in a separate window on the screen, with instant update as the cursor is moved, but also some systems allow model evaluation 'on the fly', so that a model can be checked in critical areas, giving important insights about the effects of particular combinations instantaneously. The computer tools of visualization and query can be thought of as the electronic equivalent of browsing through stacks of maps on a light table.

In terms of predictive modelling of mineral potential, there are two major lines of approach. One is to use the knowledge of experienced exploration geologists directly for modelling, the other is to use a statistical approach to calculate parameters for a model using known mineral deposits in a training area. Although the statistical approach is attractive because it is the most objective, it has the drawback that the training area used to estimate the model parameters is often inadequate. For example, the exploration may be incomplete, the known deposits are usually few in number and difficult to classify according to the correct deposit model, and the training area is not sufficiently similar to the prediction area. On the other hand, the expert system approach is attractive because deposit model criteria can be explicitly used and it appeals to many practicing geologists because it simulates their own thinking. However, the process is very subjective, the model parameters are 'guesstimates' and must be adjusted for each project area. Hybrid methods that use both statistical and expert system concepts together offer a promising compromise, as discussed in detail in Chapter 5.

## 1.6. IMPORTANCE OF DATA

The adage 'garbage in-garbage out' (GIGO), often used to describe the indiscriminate application of multivariate statistics to trashy data, certainly holds also for spatial data integration. It cannot be emphasised too strongly that the interpretation of the results of an integration study hinges strongly on the careful selection of good data.

There are numerous kinds of errors that can effect the quality of spatial data. Errors of spatial mis-registration, errors of classification, errors of measurement, errors due to poor matching of adjacent sheets, measurement errors, interpolation errors, and the list goes on. Two important types of information about data quality ideally should be associated with each dataset. The first is ***metadata***, i.e. data about data. For example, the metadata

associated with a geochemical dataset might include a detailed description of sampling methods, data preparation and analytical methods, as well as date of collection, the name of the field party, the name of the analytical laboratory, and so on. The metadata should allow any future user to be fully aware of the provenance and history of the data. The second type of information relates to the spatial variation of data quality, and is ideally a map that accompanies the 'parent' map. For example, a geochemical map of zinc variation in soil might be accompanied by a map showing uncertainty due to interpolation, such as the kriging variance. The subject of the effects of errors in data and their propagation through a GIS is relatively undeveloped, although a recent book summarizes some of the research on the subject (Goodchild and Gopal, 1989). At present, the best guard against being mislead by bad data is to get a close understanding of its limitations by display, interactive interrogation and a common sense evaluation of the results (Knox-Robinson et al., 1992).

## 1.7. DOCUMENT OUTLINE

Chapter 2 discusses a number of aspects of spatially distributed data, including data capture, data representation, and its management in data systems. Visualization of data both on computer monitors and in hard copy are introduced in Chapter 3. Chapter 4 covers the basic methods of image processing. Chapter 5 covers some of the important aspects of integration modelling. Three of the methods, using fuzzy logic operators, weights-of-evidence calculations and logistic regression, are illustrated with an application to gold exploration in Nova Scotia. The technical requirements needed to undertake spatial data integration projects, from both the hardware/software point of view and most importantly the staffing needs, are outlined in Chapter 6. Three case studies are illustrated in Chapter 7: (1) Data integration and GIS modelling in the Northwest Territories of Canada; (2) Remote sensing and GIS in mineral exploration in central western Spain; (3) Gold prediction in NE Finland. Chapter 8 ponders the future, in terms of the impact of trends in hardware.

## REFERENCES

Brodaric, B. and Fyon, J.A., 1988, OGS Fieldlog: A micro-computer-based methodology to store, process, and display map-related data, Ontario Geological Survey Open File Report 570, p. 73.

Cox, D.P. and Singer, D.A. (Eds), 1986, Mineral deposit models, United States Geological Survey Bulletin 1693, p. 379.

Eckstrand, O.R. (Ed.), 1984, Canadian mineral deposit types: a geological synopsis, Geological Survey of Canada, Economic Geology Report, v. 36, p. 86.

Goodchild, M.F. and Gopal, S. (Eds), 1989, The Accuracy of Spatial Databases, Taylor and Francis, London, p. 290.

Knox-Robinson, C.M., 1992, Geological applications of Geographic Information Systems (GIS): Introduction, Australian Institute of Geoscientists Bulletin 12, pp. 1–2.

Knox-Robinson, C.M., Robinson, D.C. and Groves, D.I., 1992, The use of Geographic Information Systems as a gold prospectivity mapping tool, with reference to the Yilgarn Block, Western Australia: requirements and limitations, Australian Institute of Geoscientists Bulletin 12, pp. 71–82.

Pflug, R. and Harbaugh, J.W. (Eds), 1992, Computer Graphics in Geology: Three-Dimensional Computer Graphics in Modeling Geologic Structures and Simulating Geologic Processes. Lecture Notes in Earth Sciences., Springer-Verlag, Berlin, v. 12, p. 298.

McCammon, R.B., 1989, Prospector II – The Redesign of Prospector: AI Systems in Government, March 27-31, 1989, Washington, D.C., pp. 88-92.

McCammon, R.B., 1990, Prospector III – Towards a map-based expert system for regional mineral resource assessment, Statistical Applications in Earth Sciences (Eds Agterberg, F.P. and Bonham-Carter, G.F.), Geological Survey of Canada, Paper 89-9, pp. 395–404.

Raper, J. (Ed.), 1989, Three Dimensional Applications in Geographical Information Systems, Taylor and Francis, London-New York, 189 pp.

Tomlin, C.D., 1990, Geographic Information Systems and Cartographic Modelling, Prentice Hall, Englewood Cliffs, New Jersey, 249 pp.

Turner, A.K. (Ed.), 1992, Three-Dimensional Modeling with Geographic Information Systems, Kluwer Academic Publishers, Dordrecht, 443 pp.

Wadge, G. and Pearson, E., 1991, GIS for geology, Terra Nova, v. 3, pp. 93–98.

# 2. SPATIALLY DISTRIBUTED DATA

## 2.1. DATA SOURCES AND DATA CAPTURE

### 2.1.1. Earth science data

In the earth sciences, one of the basic tasks is to describe and interpret phenomena through observations in the field. Observations consist of lists of characteristics (e.g., rock type and composition), measurements (e.g., strike and dip of layers), and relationships between observations at different locations. Samples of rock or of soil are collected to be later analyzed in a laboratory. By necessity, direct observations in the field are restricted to small areas and their positions are annotated on topographic maps or on aerial photographs for later compilation into thematic maps. Aid to this task is the analysis and interpretation of aerial photographs, satellite images, geophysical and geochemical maps which provide a spatial context. Accurate geographical positioning is vital in spatial integration studies.

According to Aronoff (1989, p. 1), a geographical information system (GIS) "is designed for the collection, storage and analysis of objects and phenomena where geographical location is an important characteristics or critical to the analysis". Such a definition puts the emphasis on the analysis which can be performed on spatial datasets. A GIS is seen as a decision-making tool where geographical representation is essential. In addition, the reference to phenomena implies modelling of spatial processes. Indeed, in the earth sciences, the spatial distribution of objects and phenomena is a consequence of the processes that generated them and it can be characterized for genetic interpretations. For example, we can observe that landslides tend to occur at particular locations where the geological, morphologic, topographic and hydrologic conditions are such that erosion processes and the gravity field are likely to generate sliding. Similarly, soil acidity reaches high values in the vicinity of coal refineries and metallurgical plants due to the process of releasing acid fumes in the air, or volcanogenic massive sulphide deposits occur in the vicinity of volcanic centres due to the process of submarine volcanic activity.

By modelling the position of geological objects in space and in time, it is possible to recognize relationships that can be used to predict the location and the impact of the phenomena analyzed. Spatial distribution may be characterized by the relative position between objects and on their absolute locations. Examples of relative positioning are concepts such as: 'coinciding in position with', 'within a certain distance from', 'adjacent to', 'contained in', 'oriented in direction ... with strength.' Therefore, topology, shape, orientation and distribution are fundamental spatial concepts, although the absolute locations of objects are also critical for digital representation of spatial features. Position in space, however, is only one aspect of geoscience data. In combining spatially distributed data, a clear modelling rationale must be developed that physically and genetically relates phenomena with spatial relationships. Failure to do so, reduces spatial correspondence to an unexplained coincidence which is of little predictive utility.

Typically spatial data in the earth sciences consists of field notes associated with observation points, measurements made with field instruments (including airborne and satelliteborne instruments), features being plotted on maps and photographs (pointform, linear or areal). Also common are features from previously produced maps, tables, and digital files, e.g., geophysical or other remotely sensed data. In collecting the information, however, the earth scientist simultaneously generates spatial and non-spatial

data. The latter type of data consist of observations, which can be associated with a location but which are themselves non-spatial in character, such as a variety of sample attributes (mineral composition, texture, compactness, etc.). Spatial attributes of objects usually require different management and processing as compared to non-spatial attributes, particularly when the spatial objects are lines and areas. The elementary objects in spatial characterizations are points, segments, bounded areas or polygons, surfaces and volumes.

A distinction can be made between spatial objects with sharp boundaries, such as rivers, lakes or city limits, and spatial objects represented by continuously varying values, such as topographic elevation surfaces or gravity and aeromagnetic fields, which are represented as interpolated surfaces (or volumes where 3-dimensional data are available), e.g., ore body grade characterization by drilling. Also, in some applications where time-dependent data are available, process dynamics can be represented.

Processing non-spatial data consists of selective retrievals and of computations in which the values associated with spatial objects are considered independently of their spatial characteristics. In many instances, a spatial manipulation is performed to obtain only the non-spatial attributes of a given subset of spatial objects. The non-spatial data are stored in tables which are generally separate from, but which are linked to, the spatial objects.

Computer processing methods must manipulate the data according to the spatial and non-spatial characteristics, the data type (nominal, ordinal, interval and ratio), the temporal character, and finally, the confidence or uncertainty that is associated with the data.

### 2.1.2. Data and information

Spatial data of many different types and provenances are now available to the geoscientist, who has to cope with enormous data volumes in order to take decisions about further mineral exploration efforts, and about assessing geological hazards or environmental impacts. A realistic view of the dilemma facing the decision maker is represented by the diagram in Fig. 2.1 (after Griffiths, 1974) showing the quantification future of geoscience. The development of a multitude of new sensors has provided an enormous amount of new data which has to be decoded, deciphered and manipulated in order to be interpreted and used in the earth sciences.

At present, with the development of GISs and of fast data capture techniques, even more data is available for spatial analysis. An important question is then the following: "What should we do to the data to obtain or extract information?" For instance, how can we use spaceborne or airborne images for mapping or for setting up an exploration strategy? Raw images have to be enhanced, corrected geometrically and spectrally, and photointerpreted. In the end, the results of such interpretations will be used in the decision-making process. Digital imagery is often seen with suspicion by geologists who are uncomfortable about the poor locational accuracy of the images or of the results of automatic classification. Map products, however, are no better in accuracy, since they are the results of fieldwork and of photogrammetric work which also have a statistical accuracy, i.e., a probability of being positionally correct within a given distance from the true location! It is still uncommon to have accuracy designation accompanying topographic maps.

Since the development of systematic approaches with GISs, the need has grown to analyze map accuracy and reliability before digitization. The main question, then, is the

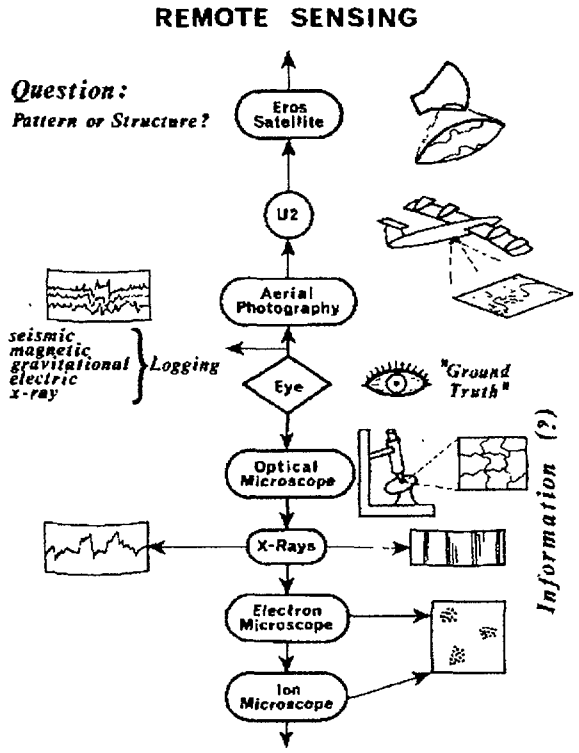



*FIG. 2.1. Proliferation of processes for data gathering (after Griffiths, 1974, p. 89, Fig. 3).*

following: "What is the information content of a map?" On any map, some objects are bound to be symbolically represented, other objects are true-to-scale, other objects are interpretations of the map maker, therefore, they are generalized in such a way that it becomes a specialist's job to capture the proper information from the map. This is particularly important if a computerized map has to be used for purposes other than the one for which it was initially made.

For example, on land property (cadastral) maps, boundaries of building blocks are absolute, so are river banks or the limits of existing public parks. In the earth sciences, however, map unit boundaries may be relative to the author's interpretation or may have a variable meaning either due to the characteristics of the map units in the legend, or to the sharpness of the contact between the units. Some contacts may represent gradual transitions, other ones may be inferred transitions, other ones may be sharp transitions and easily observable. For this reason, it becomes the earth scientist's dilemma to decide whether it is useful to digitize an existing map in a GIS or it is preferable to construct a new map from scratch. The latter option is generally ruled out due to cost!

Acknowledging that a map represents a certain degree of generalization by its author, who has introduced his interpretation skills for synthesis, it is possible to disaggregate the elementary components of a map and then recombine them for other purposes? How do we separate interpretation from factual data? How do we evaluate the degree of generalization? What is the uncertainty associated to the data in a map?

An example of a research effort to provide more usable maps for GISs is the initiative of the Italian National Geological Survey in the computerization of geological mapping for a 1:50 000 new multiuse geological cartography (CNR-SGN, 1991). A guide was published

which proposes rules and standards for transferring field data in numeric form to construct a database associated with each map and with the development of different themes (e.g., sedimentary or volcanic environment, or the understanding of the Quaternary deposits, or of geomorphologic hazards). In particular, the guide proposes to distinguish the mapping units according to their degree of objectivity in three categories:

(a) directly observable units (and those termed objective or para-objective;
(b) deduced units; and
(c) subjective and/or conventional units.

Clearly, such an approach is facilitated by using a GIS for storage and management of the data and of the associated degrees of objectivity.

Steps in the direction of institutional development for systematic capture of data in the field for geological cartography and GIS processing, have resulted in a new mapping approach at the Geological Survey of Canada (Brodaric, 1991; GSC, 1991). Data are stored in portable microcomputers during the fieldwork and all desirable elementary characteristics, including absolute location, are captured and stored for later selective representation either as symbolic, numeric or alphabetic objects. They can then be processed towards the identification of special themes. FIELDLOG is a relational database management system designed for geological field data capture (Brodaric and Fyon, 1989).

The annotation of the uncertainty associated with map data is not new in engineering geology where, for the purpose of assessing the suitability of different terrains for engineering uses, special attention is given to the transitions between mappable units (Varnes, 1974) and different weights can be associated with different transitions or contacts. Also, in mineral exploration, early applications of expert systems for 'drilling site selection' such as Prospector, allowed for different degrees of confidence or belief to be associated with map features such as faults and contacts, or geochemical anomalies (Duda et al., 1978; Duda, 1980).

Such approaches are now more readily implemented with GISs. The assessment of uncertainty associated with map data is often now a prerequisite to the construction of thematic maps.

Burrough (1986) discussed data quality in GISs, and the sources of errors and of natural variation which can affect a database and the resulting analytical work. Table 2.1 lists the possible factors controlling errors that are associated to data products in GISs.

Drummond (1987) proposed different procedures to express in probabilistic terms the quality of maps obtained within a GIS by processing data using mathematical and logical models. In her work, quality parameters for interpolated point samples and for maps with boundary lines were developed. The superposition of the results obtained by different photointerpreters in the analysis of the same material was used to compute the percentage probability of correct assignment to a bounded area. Visual representation of the results of such a probabilistic procedure is likely to be of usefulness to planners. Data quality and error propagation are two of the major research topics in GISs today.

### 2.1.3. Quantification and digitization

The previous section has discussed the transition between data and information, the latter being still data, however, with a specific meaning in operational terms so that it can be

TABLE 2.1. SOURCES OF POSSIBLE ERRORS IN GISs
(after Burrough, 1986, p. 104, Table 6.1)

| | |
|---|---|
| I. | OBVIOUS SOURCES OF ERROR |
| 1. | Age of data 13 January 1994 |
| 2. | Areal coverage - partial or complete |
| 3. | Map scale |
| 4. | Density of observation |
| 5. | Relevance |
| 6. | Format |
| 7. | Accessibility |
| 8. | Cost |
| II. | ERRORS RESULTING FROM NATURAL VARIATIONS OR FROM ORIGINAL MEASUREMENTS |
| 9. | Positional accuracy |
| 10. | Accuracy of content-qualitative and quantitative |
| 11. | Sources of variation in data: data entry or output faults, observer bias, natural variation |
| III. | ERRORS ARISING THROUGH PROCESSING |
| 12. | Numerical errors in the computer: the limitations of computer representation of numbers |
| 13. | Faults arising through topological analyses: misuse of logic problems associated with map overlay |
| 14. | Classification and generalization problems: methodology class interval definition interpolation |

used for analysis or in a decision-making process. Indeed, the fundamental task in the application of GISs is the quantification of data into a computer processable form. It is essential, before computerization to identify operational entities to be spatially located and characterized with additional non-spatial attributes. Geometrical entities with several dimensions such as points, segments, polygons, surfaces or volumes, need to be uniquely classified as sensible operational units or classes for the construction and management of a spatial database. For instance, a basic operational unit for polygons in geomorphology, is a terrain mapping unit or TMU. Each unit during field verification and from photointerpretation, has uniform characteristics while being clearly distinguishable from different adjacent units (Meijerink, 1988). TMUs can only be subdivided further using information other than the one used for their classification. As such, they represent basic entities that should be easily described and managed in a spatial database. New, broader units can be generated by the process of generalization, or smaller units can be obtained by the process of regionalization in which additional layers of spatial information are used to generate a more specific theme. TMUs are then the end result of analyses and interpretations, therefore, their information content is known and their useability when they are in computational form can be anticipated.

Similar consideration can be made for all other geometric entities in a spatial database. We can consider quantification the process of identification of such operational units, while digitization is the process of transforming such units from an analog form into

digital form. In short, "we must tell the computer what is what and what it is, spatially and topologically" (Robinove, 1989).

Digital data represent the input to the computer. As illustrated in Fig. 2.2, in a GIS we can identify the basic components as follows: (a) an input sub-system for data capture and storage; (b) a management sub-system to structure, query and handle the data; (c) an analysis and modelling sub-system to study and interpret the information; and (d) an output sub-system to render spatially and otherwise the results in interpretable and usable ways for decision making. Some spatial data may already be in digital form, such as Bouguer gravity, aeromagnetics or digital elevation, while other data might have to be digitized from original spatial representations (e.g., from maps).

There are numerous ways to capture spatial data in a computer. Typically, in the case of line data (such as boundaries of orebodies, rivers, coast lines, contacts between crystals) the lines are extracted by constructing a map for each required type of line (which is then scanned separately) as shown in Fig. 2.3(a), or the lines are extracted by human operator while the computer is recording the X–Y coordinates of the tracing stylus, as shown in Fig. 2.3(b). Various combinations and modifications of these two input methods are used.

The scanner, in contrast, produces a systematic two-dimensional raster or matrix of numbers. As it can be deduced from Fig. 2.3(a), each element of the matrix corresponds to the amount of reflected light from a small area, dA. The area dA is called a 'pixel' or picture element. Normally, the signal from the light sensor is thresholded to give a binary number of zero (0) or one (1) for each pixel: for example, 1 = sensor is located over a dark region, such as is over a line, and 0 = background. The resultant matrix of zeros and ones is usually termed a binary image or a binary picture. The data entered in this situation are intrinsically two-dimensional. Scanning can also provide greyscale or colour raster images as sets of red, green and blue components.

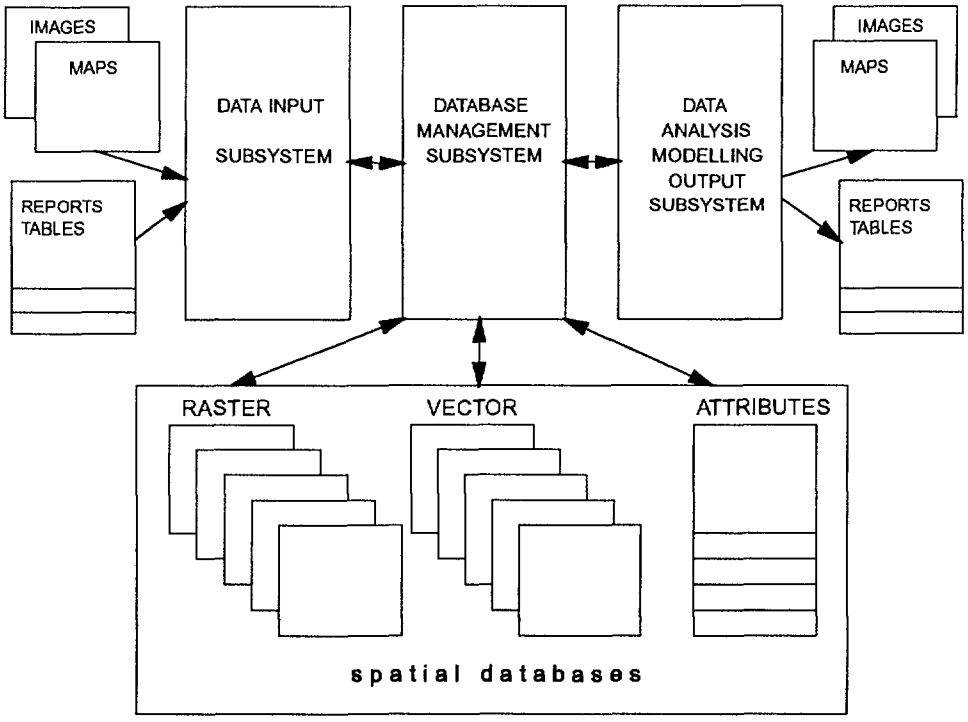



*FIG. 2.2. The basic components of a GIS.*

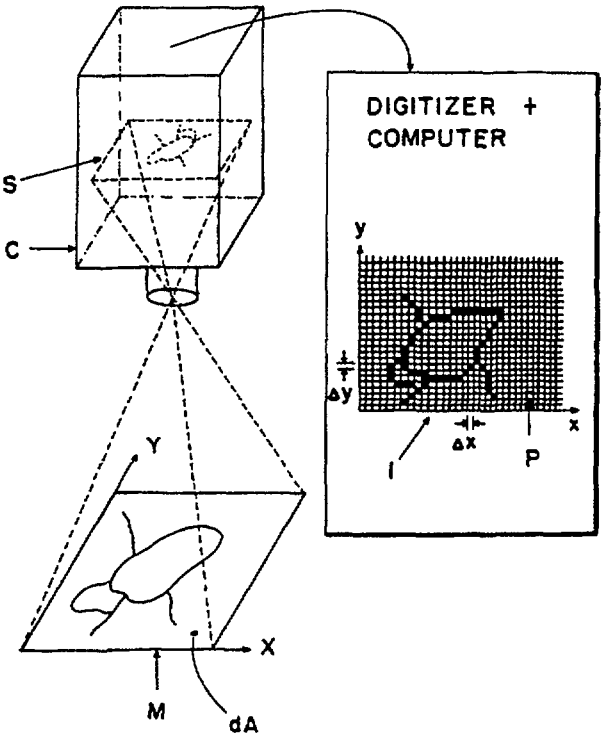



*FIG. 2.3a. A highly simplified schematic diagram of the digitizing process using a television camera or scanner. An optical image of the map M is focused onto a light-sensitive surface S in the camera C. The digitizing process quantizes the signal in space (X,Y) and amplitude, and creates a digital image which is a matrix of numbers. The numerical value of each picture element P (of I(x,y)) is proportional to the average light reflected from a small area dA on the map in the location corresponding to P.*

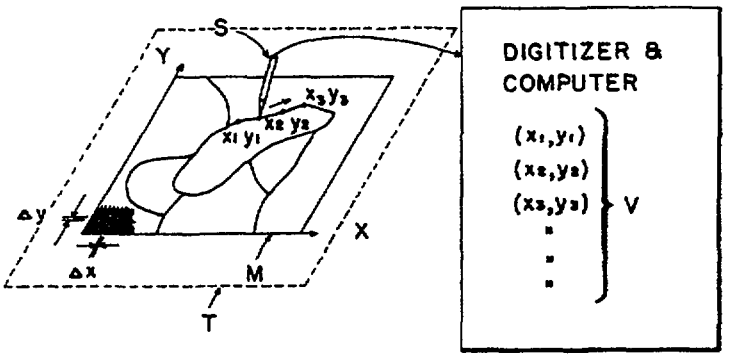



*FIG. 2.3.b. A highly simplified schematic diagram of the digitizing process using a graphic-tablet digitizer. The map M is mounted onto a tablet T. The tablet contains a grid which quantized the X and Y directions in steps of delta X and delta Y. The XY position of the stylus S is read with the spatial accuracy of delta X, delta Y. As the stylus S is moved along some trajectory on the tablet, the computer receives a string of X–Y coordinate pairs of vectors V.*

Once the visual aspect of the elementary components of spatial data has been digitized, what remains to be done is to make sure that the computer 'understands' data types, topology and attributes within an error free processable data structure. This requires checking and correcting digitizing errors, linking of point and/or segments to generate polygons, point or line interpolation to generate surfaces or volumes, and the association of tables of attributes to the individual points, line segments, polygons, etc., i.e., the construction of a spatial database. This task is generally the most time consuming stage in the construction of a database.

In some systems an interaction with the spatial data is possible for on-line interpretation (e.g., screen digitizing) or image editing for better control of the stored data and for the preparation of inputs for analysis.

## 2.2. DATA REPRESENTATION

### 2.2.1. Data structures for non-spatial data

Database management systems, or DBMSs, are computer programs for storing, manipulating and retrieving items from collections of data files stored in a structured fashion and termed a database. In a database, the interrelationships coded between different sets of data are used for processing and retrieving. In general, a DBMS assumes the existence of many users, and the absence of unnecessary redundancy which would complicate management and updating. While the physical description of a database deals with the location of the various parts of the database in the computer, it is the logical design that represents the user's view or conceptual model of the relationships between the datasets stored in the database.

A common approach in database design is the entity-relationship model of Chen (1976). In order to illustrate this model, suppose that we have constructed a database of mineral deposits for an area of interest. A deposit entity set represents the generic structure of the object or entity described. Each deposit represents an entity which has a number of attributes (e.g., location, name, type of mineralization) each with a range of possible values, such as UTM coordinates for location.

We can also define relationship sets, such as deposit ownership, mineral commodities present, or association of host rocks to the ore, which allow the creation subsets and retrievals. Specific relationships between entity sets are classified as one-to-one, one-to-many, and many-to-many mappings. For instance, one company may own one deposit, another company might own more than one deposit, or each deposit may produce several commodities which are present in several mineral assemblages.

Figure 2.4 illustrates one example of an entity-relationship model for a mine owner database. Table 2.2 lists some of the data extracted from Roscoe (1984) who described several mineral occurrences in a frontier area of the Northwest Territory of Canada. In its present form, Table 2.2 represents a relational database from which it is possible to extract information by performing operations between tables (termed relational joins). For instance, a relational operation is possible between the mine owner's table and the occurrence table to associate owner names and occurrence names, or between the commodity table and the occurrence table to associate mapping symbols to latitude and longitude on a topographic map.

A relational database contains relationships between entities directly represented as tables, as it was done for the Occurrences–Commodities table in Table 2.2. The simplicity of such a representation is evident.

The requirements for a true relational database are that the tables have the following properties:

(i) rows cannot be duplicated (i.e., an entity set cannot have rows whose entire contents of values are identical);
(ii) each row must be unique and is identified by a 'key field' , column or columns;
(iii) no key field column(s) can have null values.

Also, a relational database must contain a minimum amount of data redundancy to allow the join operations. Codd's (1970) theory of 'normal form' establishes the rules of relational database construction. Recent progress in software development have now made this database structure sufficiently efficient for general widespread use. Another type of data structure is the hierarchical data structure system in which a given entity set is the root of a tree and a set of parent-child pointers allow linkages from level to level of the hierarchy. Figure 2.5 shows this tree structure for the example of mine owner database in Table 2.2. As can be seen in the illustration, such a structure requires the repetition of attributes for each entity. While this structure is easier to construct, and to handle one-to-many relationships, the many-to-many relationships (which cannot exist in the hierarchical data structure) are much more cumbersome to handle as exceptions. In the relational database structure many-to-many relationships are avoided altogether by the construction of several one-to-many relationship tables. Other types of DBMS structures have been developed in the past or are still being used in some systems (Healey, 1991), however,

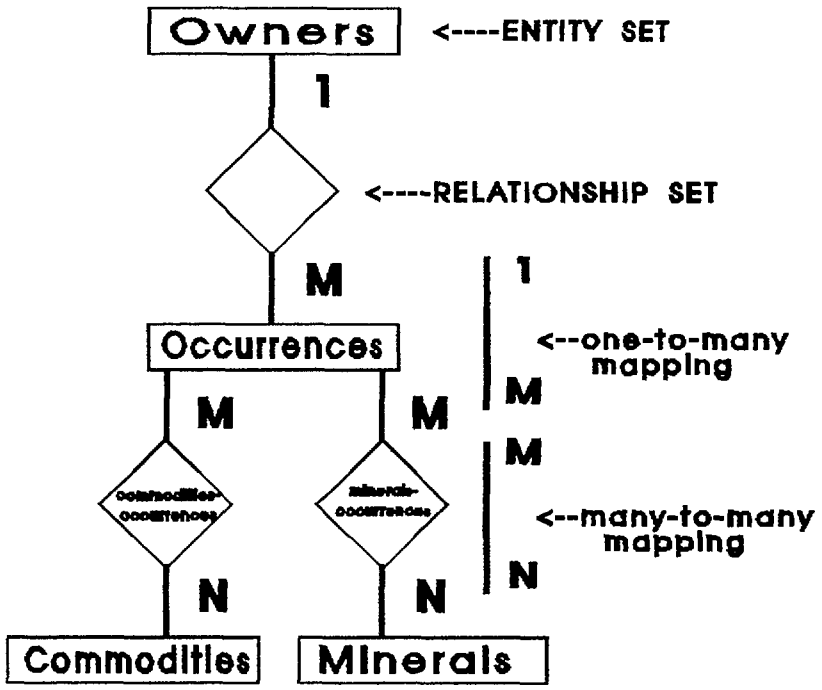



*FIG. 2.4. The entity-relationship model for the mine owner database in Table 2.2.*

relational databases seem to be the most widely adopted at present. The relational data structure is also common for spatial data. For instance, the column Polygon in Fig. 2.10, links the mineral Dep._# and Name to a graphic database in which a geological map has been stored.

TABLE 2.2. THE STRUCTURE OF THE MINE OWNERS RELATIONAL DATABASE

The Mine Owners Table

| Owner_# | Name | Year_of_activity |
|---|---|---|
| 1 | Cominco | 1976 |
| 2 | Trans-Canada | 1968 |
| 3 | Roberts | 1964 |
| 4 | Noranda | 1977 |

The Occurrences Table

| Occ._# | File_# | Name | Lat. | Long. | Owner_# | NTS_# |
|---|---|---|---|---|---|---|
| 1 | 061589 | POMIE | 66 10′ | 107 02′ | 1 | 76J/3 |
| 2 | 502678 | COT | 66 42′ 10" | 107 26′ 40" | 2 | 76J/11 |
| 3 | 502693 | PISTOL L. | 67 03′ | 108 47′ | 3 | 76N/2 |
| 4 | 506486 | TURNER L. | 67 14′ | 108 57′ 30" | 3 | 76N/2 |
| 5 | 080737 | YON | 66 35′ | 107 28′ | 4 | 76J/11 |
| 6 | 061590 | JWC | 66 35′ | 108 00′ | 1 | 76J/12 |

The Commodities Table

| Comm._# | Name | Symbol |
|---|---|---|
| 1 | gold | Au |
| 2 | copper | Cu |
| 3 | uranium | U |
| 4 | nickel | Ni |
| 5 | cobalt | Co |

The Minerals Table

| Min._# | Name |
|---|---|
| 1 | uraninite |
| 2 | pyrite |
| 3 | chalcopyrite |
| 4 | pitchblende |
| 5 | bornite |
| 6 | malachite |
| 7 | hematite |
| 8 | arsenopyrite |
| 9 | pyrrotite |
| 10 | niccolite |
| 11 | sofflorite |
| 12 | gersdoffite |
| 13 | galena |
| 14 | gold |

TABLE 2.2. (cont.)

The Occurrences_Commodities Table

| Occ._# | Comm._# |
|---|---|
| 1 | 3 |
| 2 | 1 |
| 2 | 2 |
| 3 | 1 |
| 4 | 4 |
| 4 | 2 |
| 4 | 6 |
| 4 | 5 |
| 4 | 1 |
| 5 | 3 |
| 5 | 2 |

The Occurrences_Minerals Table

| Occ._# | Min._# |
|---|---|
| 1 | 1 |
| 1 | 2 |
| 1 | 3 |
| 2 | 2 |
| 2 | 3 |
| 2 | 8 |
| 2 | 9 |
| 3 | 2 |
| 3 | 8 |
| 3 | 9 |
| 4 | 2 |
| 4 | 3 |
| 4 | 8 |
| 4 | 9 |
| 4 | 0 |
| 4 | 11 |
| 4 | 12 |
| 4 | 13 |
| 4 | 14 |
| 4 | 15 |
| 5 | 1 |
| 5 | 3 |
| 5 | 4 |
| 5 | 5 |
| 5 | 6 |
| 5 | 7 |

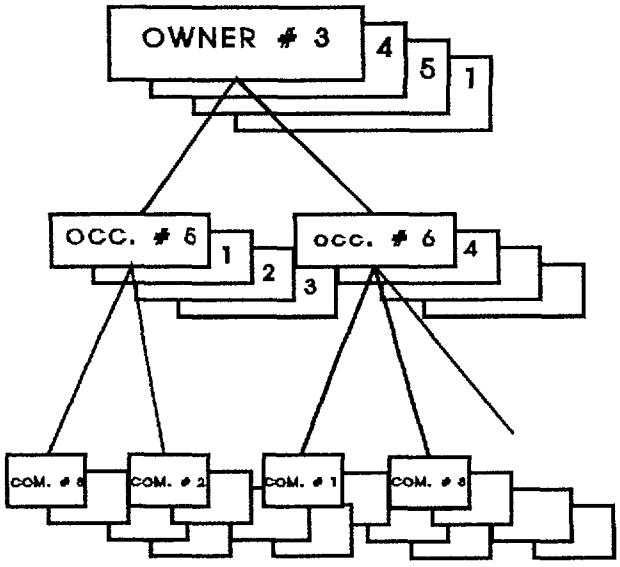



*FIG. 2.5. The tree structure of the hierarchical model for the mine owner database in Table 2.2.*

### 2.2.2. Spatial data structures

This section describes the most common data structure used in GISs. The user's perception of a phenomenon that actually exists or the human conceptualization of reality has to be represented and stored in a complete database. There are several ways of organizing these geographical data in a computer. The data model of the earth or the description of entities and their relationships are represented in terms of basic entities such as points, lines, areas and surfaces. The non-spatial data, or the sets of attributes describing the scales of properties that apply to the entity, are stored usually in relational database management systems. The spatial distributions of points, lines, areas and surfaces are represented in digital form in two basic types of spatial models: tessellation and vector models. Vector models have the line as the basic logical units in a geographical context. A series of X–Y pointer locations along the line are recorded as the components of a single data record. Points are recorded as lines of zero length, areas or polygons constitute lines with common beginning and ending points. Tessellation or raster models have as the basic data unit, a unit of space for which entity information is explicitly recorded. The most common vector models are the whole polygon structure and the topologic model.

In the whole polygon structure, or spaghetti model, shown in Fig. 2.6, each polygon is encoded in the database as one logical record and it is defined by a string of X–Y coordinates representing a closed area. This model is often used for applications that are limited to the simple forms of computer-assisted cartographic production. In the topologic model, shown in Fig. 2.7, the spatial relationships among entities are explicitly recorded. The basic logical entity is a straight line segment that begins or ends at the intersection with another line or at a bend in the line. Each line segment is recorded with the coordinates of its two end points. In addition, the identifiers, or names of the polygons on either side of the line are recorded, thus the more elementary spatial relationships are explicitly retained and can be used for analysis. Moreover, the basic data are stored in a non-redundant manner.

The main encoding schemes for the topologic model are:

(1) The Dual Independent Map Encoding (DIME) structure was devised by the US General Bureau to store digitally urban areas with topologic information for demographic analysis (US Department of Commerce, 1990). In the DIME file, each line segment is spatially defined according to the definitions of the model, using both street addresses and UTM (Universal Transverse Mercator) coordinates. The main disadvantage of this model is that since segments do not occur in any predefined sequence, the retrieval of any particular segment requires a sequential, exhaustive search on the entire file.

(2) POLYVERT is an Arc-Node structure implemented at the Harvard Laboratory for Computer Graphics (Peucker and Christman, 1975). This structure overcomes the major retrieval difficulties by storing each type of data entity separately in a hierarchical data structure. The basic entity is the 'arc', a sequence of straight lines which begins and ends at a 'node'. A node is the intersection point between two arcs. Polygons are areas that are completely bounded by a set of arcs. The hierarchical structure allows selective retrieval of only specific classes of data at atime. Various attributes can be easily included and linked to the geometry.

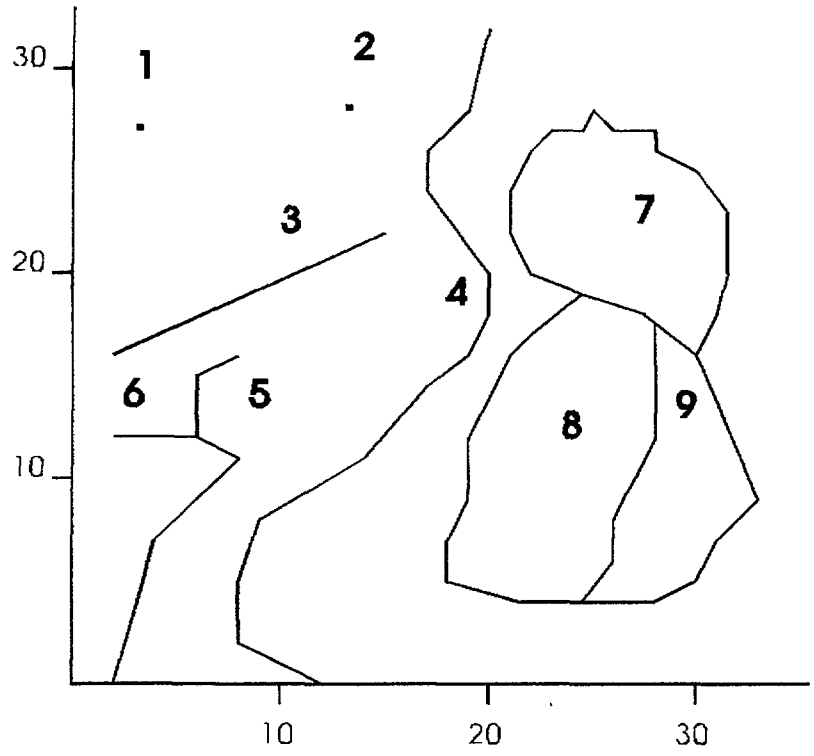



*FIG. 2.6. Example of the whole polygon or 'spaghetti' data model. See data structure in Table 2.3. Some authors use this term for vector data that is devoid of attributes, except (x,y) coordinates.*

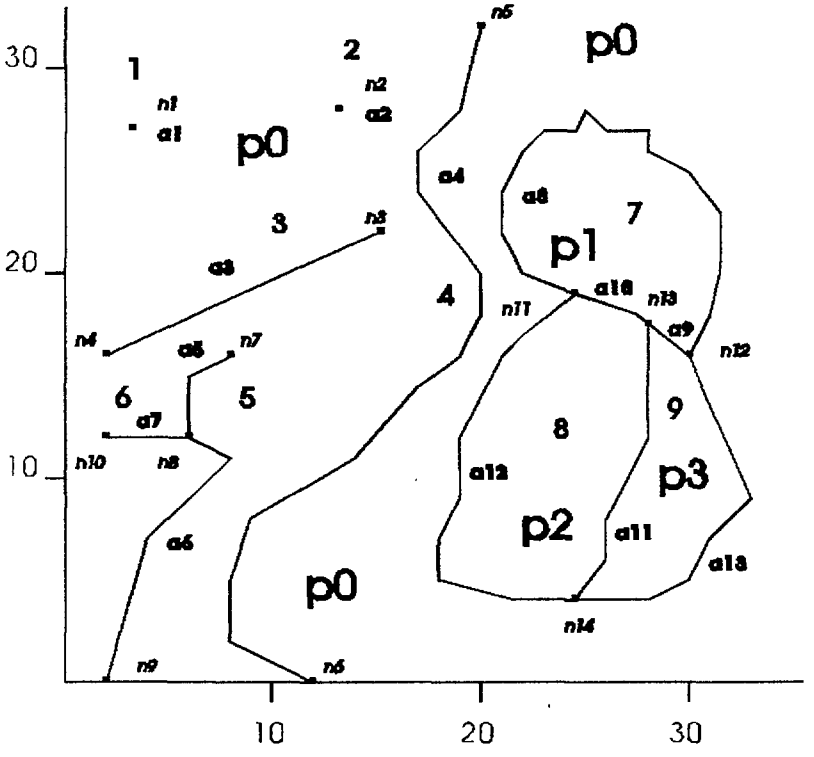



*FIG. 2.7. Example of the topological data model. See data structure in Table 2.4.*

TABLE 2.3. STRUCTURE OF WHOLE POLYGON DATA MODEL FOR FIG. 2.6

| FEATURE | NUMBER | NAME | LOCATION (X-Y) |
|---|---|---|---|
| Point | 1 | drillhole1 | 3,27 |
| Point | 2 | drillhole2 | 13,28 |
| Line | 3 | fault | 2,16 14,22 |
| Line | 4 | road | 11,0 7,2 7,5 8,8 13,11 16,14 19,16 20,18 20,20 17,24 17,26 19,28 20,32 |
| Line | 5 | river | 2,0 4,7 8,11 6,12 6,15 8,16 |
| Line | 6 | creek | 2,12 6,12 |
| Polygon | 7 | rockunit3 | 29,16 30,18 31,20 31,23 30,25 28,26 28,27 26,27 25,28 24,27 23,27 22,26 21,24 21,22 22,20 24,19 26,18 29,16 |
| Polygon | 8 | rockunit2 | 24,19 22,17 21,26 20,14 19,12 19,9 18,7 18,5 21,4 23,4 24,6 24,8 25,10 26,12 26,16 26,18 24,19 |
| Polygon | 9 | rockunit1 | 26,18 29,16 30,13 31,9 29,7 28,5 26,4 23,24 24,6 24,8 25,10 26,12 26,16 26,18 |

TABLE 2.4. STRUCTURE OF TOPOLOGIC DATA MODEL FOR FIG. 2.7

| NODE | ARCS |
|---|---|
| n1 | a1 |
| n2 | a2 |
| n3 | a3 |
| n4 | a3 |
| n5 | a4 |
| n6 | a4 |
| n7 | a5 |
| n8 | a5 a6 a7 |
| n9 | a6 |
| n10 | a7 |
| n11 | a8 a10 a12 |
| n12 | a8 a9 a13 |
| n13 | a9 a10 a11 |
| n14 | a11 a12 a13 |

| ARCS | NODE FROM | NODE TO | RIGHT POLY. | LEFT POLY. |
|---|---|---|---|---|
| a1 | n1 | n1 | p0 | p0 |
| a2 | n2 | n2 | p0 | p0 |
| a3 | n4 | n3 | p0 | p0 |
| a4 | n6 | n5 | p0 | p0 |
| a5 | n8 | n7 | p0 | p0 |
| a6 | n9 | n8 | p0 | p0 |
| a7 | n10 | n8 | p0 | p0 |
| a8 | n11 | n12 | p1 | p0 |
| a9 | n12 | n13 | p1 | p3 |
| a10 | n13 | n11 | p1 | p2 |
| a11 | n14 | n13 | p3 | p2 |
| a12 | n14 | n11 | p2 | p0 |
| a13 | n12 | n14 | p0 | p3 |

| POLYGON END | ARCS |
|---|---|
| p1 | a8 a9 a10 |
| p2 | a10 a11 a12 |
| p3 | a9 a13 a11 |

| ARCS | START | COORDINATES (intermediate) | |
|---|---|---|---|
| a1 | 3,27 | | 3,27 |
| a2 | 13,28 | | 13,28 |
| a3 | 2,16 | | 14,22 |
| a4 | 11,0 | 7,2 7,5 8,8 13,11 16,14 19,16 20,18 20,20 17,24 17,26 19,28 | 20,32 |
| a5 | 6,12 | 6,15 | 8,16 |
| a6 | 2,0 | 4,7 8,12 | 6,13 |
| a7 | 2,12 | | 6,12 |
| a8 | 24,19 | 22,20 21,22 21,24 22,26 23,27 24,27 25,28 26,27 28,27 28,26 30,25 31,23 31,20 30,18 | 29,16 |
| a9 | 29,16 | | 26,18 |
| a10 | 26,18 | | 24,19 |
| a11 | 23,4 | 24,6 24,8 25,10 26,12 26,16 | 26,18 |
| a12 | 23,4 | 21,4 18,5 18,7 19,9 19,12 20,14 21,16 22,17 | 24,19 |
| a13 | 29,16 | 30,13 31,9 29,7 28,5 26,4 | 23,4 |

(3) The relational structure, where the topologic data are organized similarly to the arc-node model, but the non-spatial attribute values are stored in relational tables using relational database management systems. The attribute data are stored in tables where the columns are the different fields or attributes, and one table includes the points or nodes of the spatial database. Several commercial general purpose relational database management systems are available in the market.

(4) The Digital Line Graph structure developed by the US Geological Survey. It is used to compile the USGS topographic maps (7.5 and 15 minute series). The data structure of these files is subdivided into different thematic layers. One consists of boundary information, a second of hydrographic features, a third has a transportation network and a fourth is based on the Public Land Survey System.

The main tessellation models are as follows: (a) grid, (b) nested, and (c) irregular models.

(a) In the grid and other regular tessellations, square, triangular, and hexagonal meshes have been used. The most widely used is the square grid. Figure 2.8 shows the rasterization of the three polygons in Fig. 2.7. In this raster array, each cell is referred by a row and a column number and a number representing the value of the attribute being mapped. The origin of the raster is usually the upper left corner, but is sometimes the lower left. One of the limitations of a square grid is that the distance between diagonal neighbours is not at the same distance as between the neighbours in the four cardinal directions. In a hexagonal mesh all neighbouring cells of a given cell are equidistant from the centerpoint of that cell. Radial symmetry makes this model advantageous for radial search and retrieval functions. Triangular tessellations have the disadvantage that not all triangles have the same orientation, making several procedures involving single cell comparisons very complex. However, this characteristic gives triangulation structures a big advantage in representing terrain and other typical surface data. Hexagonal and triangular tessellations have the disadvantages that they cannot be recursively subdivided into smaller cells of the same shape as the original cells and their numbering system is more complex than that of a square model. In terms of processing efficiency in general procedures to complete spatial properties such as area and centroid calculations, or to perform spatial manipulations such as overlays and windowing, the algorithms initially devised for square grids can easily be modified to work in the case of triangular or hexagonal systems. These in fact have the same order of computational complexity (Ahuja, 1983).

(b) In the nested tessellation models, cells are recursively subdivided into smaller cells with the same shape and orientation. the most studied model is the quadtree based on the recursive decomposition of a grid (Peuquet, 1984). As shown in Fig. 2.9, a quadtree first encloses the area in consideration within a square, and subdivides the square into four subquadrants. Each quadrant is then recursively subdivided into four subquadrants until all of them are uniform with respect to the image value, or until a predefined resolution level is reached. With their variable resolution and natural subdivisions into hierarchical patches, quadtrees are ideal for handling large geographical areas (Mark and Lauzon, 1985). A disadvantage of quadtree models is that they appear are not invariant to translation, rotation or scaling; shape analysis and pattern recognition are difficult and need intermediate procedures.

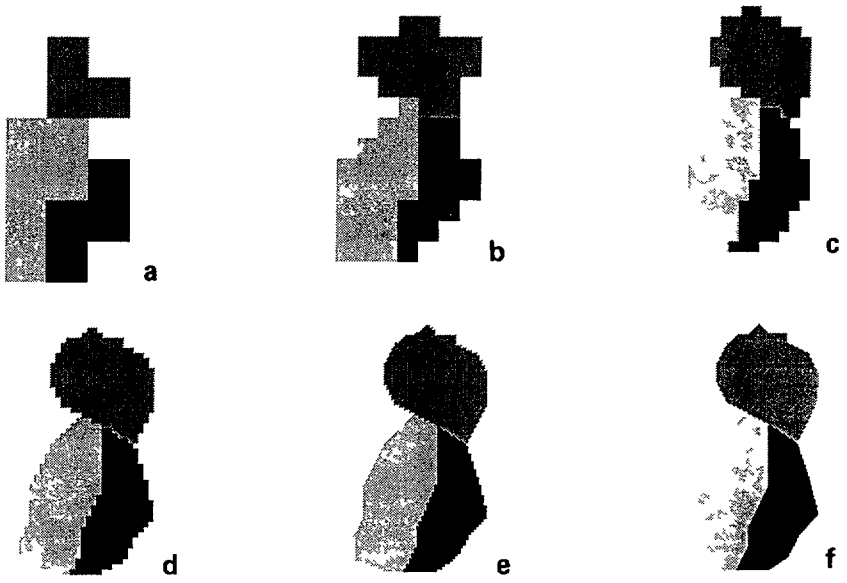



*FIG. 2.8. Rasterization of the three polygons in Fig. 2.4, using different raster resolutions of (a) 8 × 8, (b) 16 × 16, (c) 32 × 32, (d) 64 × 64, (e) 128 × 128 and (f) 256 × 256 cells to generate the same numbers of pixels. Note the blockiness of the lower resolutions and the poor rendering of the polygons.*

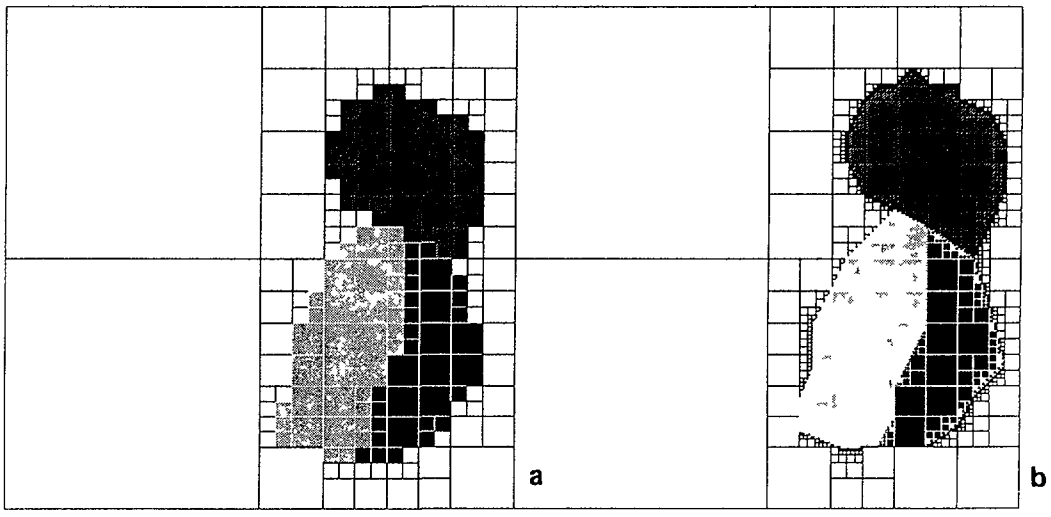



*FIG. 2.9. Quadtree representation of two of the images in (c) and (f) of Fig. 2.8, the 32 × 32 and the 256 × 256 pixel raster images in (a) and (b), respectively.*

(c) An irregular tessellation consists of an irregular net of interlocking triangles that can be adjusted to reflect the density of data occurrences within an area. In dense areas the cells are small while in sparse areas cells are large. The Triangulated Irregular Network (TIN) is composed of Delaunay triangles, where each node of the mesh has an elevation value. TINs are used to represent terrain data for digital elevation models, for hill shading and other land surface representations A TIN is based on irregularly distributed points that are first triangulated into a series of connected facets. The triangulation algorithm allows the generalization of the surface by

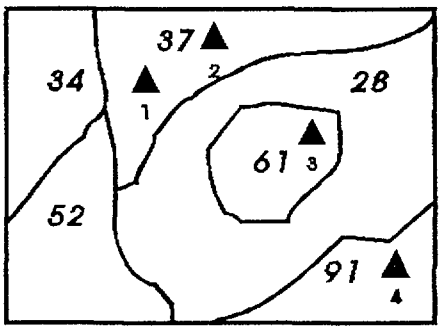
selecting points based on Chebyshev's approximation (Poiker and Griswold, 1985). Thiessen polygons or Voronoi diagrams, also termed Dirichlet tessellations, constitute other irregular models where the polygons are convex and have a variable number of sides. They are used efficiently in analysis of adjacency and of proximity. Irregular tessellations are difficult to generate and not well suited for a number of spatial analyses, such as overlaying procedures (Peuquet, 1984).

The majority of the current applications in both image processing and GISs can be handled by using raster data, however, the complete functions of a spatial data processor require both raster and vector data types. Ideally, a system should have a data structure independence, i.e., the user should be insulated from the data structure when performing any given application. The selection of the data structure to be used should be done by the system, automatically choosing either vector or raster, according to a decision on performance, both speed and accuracy.

### 2.2.3. Relationships between spatial and non-spatial data structures

The similarity in structure between the spatial data in Table 2.3 and the non-spatial data in Table 2.1 is evident. Generally, in processing the spatial data, the linkage is used between the two types of data. This linkage is represented by the polygon identification entry in the attribute table. This allows the polygon data to be processed after consulting the attribute table or to process the attribute data after consulting the polygon table.

In raster data models, a similar simple procedure is followed, because the polygon identifier is stored as a pixel value, while polygon attributes are stored in associated tables. Figure 2.10 provides a simplified example of association between spatial and non-spatial data. In most GISs the spatial and the non-spatial databases are stored separately, however, hybrid and integrated data models have been designed for large datasets. In the hybrid data models the spatial data are stored on very fast devices, while the attribute data are stored in commercial relational database systems. In integrated GIS data models, both spatial and attribute data are stored together in relational tables.




Mineral Deposit Table

| Dep._# | Name | Polygon_# |
|---|---|---|
| 1 | Oro | 37 |
| 2 | Silver city | 37 |
| 3 | Plomblere | 61 |
| 4 | Alberta | 91 |

*FIG. 2.10. Schematic example of spatial data and of associated attribute data. The polygon identifier in the table represents the linkage between the two data types.*

### 2.2.4. Data representation for integration modelling

A general approach in GIS analysis is to construct a spatial-physical rationale for guiding the analysis of multiple data layers and their connected tables over a study area. For example, Green and Craig (1984) identify one of the critical problems in mineral exploration which is to establish a connection between geological, geophysical or geochemical observations at the earth's surface and mineralization at depth. The complexity of the problem is, of course, greater when dealing with multiple datasets where geology and remotely sensed data are integrated with geophysics and geochemistry, for instance to obtain a multivariate classification.

In regional exploration programs, a physical model can connect geophysical data to physical rock properties, therefore enabling the prediction of such properties that can be compared with the spatial information provided by a geological map. The study of the difference between the predicted and the observed spatial pattern of the rocks is bound to indicate potential exploration targets. According to Green and Craig (1984), models of increasing complexity in such exploration strategy are:

(i) A Zeroth Order Model (assuming that a given rock type, when outcropping it produces the same sensor response);
(ii) Proportional Mixing Models (sensor is influenced by a mixture of rocks);
(iii) Distance-Weighted Mixing Models (sensor is influenced by adjacent rock masses more than by distant ones); and
(iv) Complex Models with Ancillary Data (using additional datasets, for example integrating vegetation-distribution to explain the contribution of pixel brightness in Landsat data, or a digital elevation model to account for gravity anomalies).

The limited success in the application of multivariate pattern-recognition techniques to obtain more useful geological maps is probably a consequence of the lack of sound geological and geophysical rationales in which the models are adequate to predict natural complexity. In this area much research is now necessary.

Table 2.1 shows a list of the sources of possible errors in GISs. How can we represent errors in a GIS? To answer this question, we can imagine a computer representation of a map in which each polygon is defined by a set of fuzzy arcs or segments, each consisting of error bands within which the positioning of objects has greater uncertainty or ambiguity than outside. We can also analyze the nature of map unit boundaries and characterize their often statistical nature, as done by Burrough (1986) who distinguished abrupt boundaries, those dividing a trend, and the ones resulting from sampling variations. Indeed, geological observations leading to a map, are associated to an inherent uncertainty which is generally recorded in fieldnotes and seldom transferred to a final cartographic representation. This is mostly due to mapping tradition and to the unavoidable simplification in map generalization (and in spatial interpolation as well).

The recent developments of GISs now allow to capture, store, represent and process such fuzzy objects using the concepts of fuzzy sets. A membership function in fuzzy set theory (Zadeh, 1965; Kaufman 1975) is used to represent the degree of membership of a cell (pixel, polygon, sub-polygon, segment, or point) to a set. According to Burrough (1986) such a function can be used to model the vagueness of map unit 'impurities' in the combination of attributes in map overlaying processes for the construction of thematic maps for resource assessment or mineral exploration.

A generalized representation for the integration of geoscience data in GISs was proposed by Chung and Fabbri (1993) who considered most maps, and enhanced or classified images after photointerpretation, as the end results of delicate decision processes. In mineral exploration, for instance, only such results should be used to obtain favourability functions with respect to a given exploration target. If we consider our results in the form of M layers of sets of map units in a study area where a given exploration target is sought, a proposition can be formulated for each pixel in the study area, "that the pixel contains a mineral deposit of a given type."

The probability of either continuous or non-continuous observations can be quantized, i.e., expressed in convenient intervals which can be regarded as evidences of the proposition. A subsequent transformation can be computed to obtain a relative favourability index function for each layer which provides measurements related to the 'sureness' that the proposition is true given the evidence at the pixel in each layer. Such sureness can be interpreted in various ways: as a probability, a certainty, a belief, a plausibility, possibility or compatibility, and further combined in a compound function which integrates the favourability functions from all the m layers. Sureness can range between 0 (low) and 1 (high). Those authors provided different interpretations of the favourability function as: (1) conditional probability, (2) certainty factor (Shortliffe and Buchanan, 1975), (3) Dempster-Shafer belief function (Shafer, 1976), and (4) fuzzy logic (Zadeh, 1965).

They considered some methods of estimation for such interpretations in three separate situations in mineral exploration: (a) there are few or no recorded occurrences of mineralization existing in environments similar to that of the study area; (b) there are known occurrences in the study area; and (c) there are known occurrences only outside the study area.

Chung and Fabbri (1993) defined data representation as a transformation of an experience of the real world into a computational domain. As such, it must comply with models and rules to provide us with useful information. For this reason, the quantitative representation of spatially-distributed map patterns or phenomena plays a pivotal role in integration because it determines the type of combination rules applied to them. The difficulties of understanding the limitations (both application dependent and probabilistic) associated with the construction of favourability functions have to be resolved to obtain useful thematic maps for exploration, for geological hazard assessment, or for environmental impact assessment. A GIS is an ideal tool for this development.

## 2.3. INTERACTION WITH THE DATA

The capability to easily interact with the data is at the root of any database management system (DBMS). In GISs, however, spatial databases require interaction with spatial objects and some spatial functions are outside the standards of conventional query languages such as the Structured Querying Language (SQL) for relational database systems (ANSI, 1989). SQL is a database interface (a language) for applications which can be expressed in terms of tables. Spatial data can be mapped into such tables, however, this type of mapping or representation leads to unnecessarily complex queries.

For instance, overlays which require a spatial operation between data layers, are not directly computable within the conventional set operators of SQL. Egenhofer (1992) discussed non-standard extensions of SQL to incorporate geometrical and pictorial spatial

operators. According to him, they seem to burden the language making retrievals overly complex and inefficient. Other solutions, beside spatial extensions of SQL, are to process the queries to convert the spatial operators into more standard queries, or to disassemble the queries into more manageable components to retrieve the data which are later processed using special-purpose user-supplied programs to perform the GIS operations.

SQL is based on 5 relational operators, termed 'selection', 'projection', 'Cartesian product', 'set union' and 'set difference', in addition to a few extensions for operations on 'tuples' (unique combinations of values in rows of relational tables), such as 'aggregate functions' and 'views'. The SQL syntax is based on a clause termed **SELECT-FROM-WHERE**. For example, to retrieve the mine owners of all gold occurrences in NTS 76N/2 in Table 2.2, the following query can be formulated:

```
SELECT   occurrences.owner_#, occurrences.name, occurrences.file_#
FROM     occurrences, commodities, occurrences_commodities
WHERE    commodities.name = 'gold' and occurrences.NTS_# = '76N/2' and
         occurrences.occ_# = occurrences_commodities.occ_#
```

The result of this query can be the display of a table such as:

| Occ._# | Owner_# | Name | File_# |
|---|---|---|---|
| 3 | 3 | PISTOL L. | 502693 |
| 3 | 4 | TURNER L. | 506486 |

An introduction to the structural query language can be found in Korth and Silberschat (1986). Beside the management of spatial data and of the associated attribute data, GIS processing uses a variety of analysis functions. Aronoff (1989) grouped these functions as follows:

(1) retrieval/classification/measurement (spatial data and attributes are retrieved but only the attribute data are modified or created);
(2) overlay operations (between data layers);
(3) neighbourhood operators (to evaluate the characteristics of the area surrounding a speecified location);
(4) connectivity functions (to accumulate values over the area being traversed).

In many GISs, special command languages or algebras facilitate the interaction for analysis and modelling tasks. Table 2.5 lists several modelling expressions from a raster-based GIS which has built-in capabilities of a relational database and of an image processing system.

Modelling algebras can vary considerably from system to system and can make use of keyboard entries, of mouse selectable menus or of icon assisted navigation tools, including facilities for generating more powerful commands and ad hoc icons, i.e., special purpose algebra using available commands or programming new commands. Some systems allow a sequence of operations to be recorded as a batch file, that can be edited at will, assigned a new command name and used as a new function or a new macroinstruction. Other systems make use of interactive icon-assisted environments such as WINDOWS. The main purpose of modelling algebra is to facilitate the generation of more complex processing tasks.

TABLE 2.5. INTERACTIVE EXPRESSIONS FROM THE MODELLING LANGUAGE OF ILWIS (Valenzuela, 1988; Gorte et al., 1988; ITC, 1992), A GIS WITH EXTENSIVE ANALYTICAL TOOLS

(1) MAP2:=TABLE.COLUMN[MAP1]

Reclassification of a raster map, MAP1, according to attribute values in column COLUMN of table TABLE, associated to MAP1.

(2) LAI:=((band7-band5)/(band7+band5))+1*127

Store on disk the raster map LAI (leaf area index) resulting from subtracting band5 from band7, dividing their difference by their sum, adding the value 1, and multiplying the results by 127. Band7 and band5 are the names of the input raster images.

(3) MAP_C:=IF(MAP_A>20,MAP_B+3,MAP_B-2)

IF the values in MAP_A are larger than 20 THEN add 3 to MAP_B, ELSE subtract 2 from MAP_B.

(4) SLOPE()=SQRT(@1*@1+@2*@2)*100

Create function SLOPE. @1 and @2 indicate parameters to be substituted by image names during interaction. Here they represent the height difference (gradient) per pixel in the X resp. Y direction.

SL_MAP:=SLOPE(MAP1,MAP2)

MAP1 and MAP2 are the two variables for @1 and @2.

(5) MAP2:=NBMIN(MAP1#)

NBMIN is a neighbourhood aggregation function which finds the minimum value. #[] is a select neighbourhood operator for a 3 × 3 window where the neighbor positions are:

1 2 3
4 5 6
7 8 9

# alone indicates the entire 3 × 3 window.

(6) ITER:MAP:=NBMIN(MAP1#+NBDIS)

Calculate a distance map from an initial map, MAP1. Stops only if map MAP does not change any more. ITER is the iteration function and NBDIS is a distance filter which uses the following constant values:

7 5 7
5 0 5
7 5 7

## 2.4. DATA SYSTEMS AND DATA TRANSFER

### 2.4.1. Overview of data systems

An important trend in the geosciences is the clear move towards microcomputers and workstations as GIS platforms. This guidebook is concerned with low-cost analytical systems to be used as decision tools in resource exploration or in environmental applications. More expensive systems designed for cartographic production are either peripheral to such applications or represent final platforms on which to transfer the information once the decisional strategy has taken a stable form. Table 2.6 depicts the main capabilities of some selected PC-based systems. The criterion for the selection of the systems listed in Table 2.6 was the availability of the systems to large numbers of users and the variety of different characteristics in their design. It should be noted that the list is by no means complete and does not try to identify the 'better' systems.

The development of GISs is dynamic and there are several tens of systems available on the market. Such systems as GISs, are still far from being subjected to satisfactory evaluations because even very complete benchmarks do not seem to reveal all the idiosyncrasies and weak and strong points of systems that might appear under a real world use. In the table we can see that PC-DOS is the most common operating system and that few GISs use also OS2 or the MacIntosh operating system. Most analytical systems, even if they can handle and process vectorial data structures in a limited way, make provision for vector-to-raster conversions for the analytical process. All systems in the table have extensive digitizing capabilities, however, only a few possess an internal database management system. Those which do not, generally rely on external DBMS and have provision for table import/export from/to some of the most common systems (e.g., ORACLE or dBase IV). In particular, some systems are hybrid GISs and image processing systems (e.g., IDRISI, ILWIS), or are mainly image processing systems (ERDAS, EASI/PACE) or GISs (PC-Arc/Info, SPANS, PAMAP) or computer aided design and drafting systems (AutoCAD). The selection of a particular system for given tasks should be based on whether their design fits the desired tasks. Some systems such as IDRISI and ILWIS were designed within academic environments as teaching tools, and for that reason they are particularly simple to use. Prices for PC-based GISs range from a few hundred to a few tens of thousands of dollars.

### 2.4.2. Spatial data transfer

As we have seen in Section 2.2, several spatial data structures exist which imply different topologic encoding or different tessellation models. Gutpill (1991) discussed different levels of abstraction in a spatial database as follows: (a) the real world as it exists; (b) the conceptual data model relevant to a specific need but independent of data structures; (c) the logical data model which defines data organization (data structure); and (d) the physical data model which specifies machine implementation of data structure (file structure). The implication in those abstraction levels is that a spatial database is the result of modelling reality by constructing a knowledge base where conceptual and logical data models are represented in a specific physical model. Such a process of abstraction and of representation should then be reversed in order to transfer the data from one system to another using different physical data models. The conceptual data model in the source system should be well understood so that the appropriate data structure can be rebuilt in the target system. Data exchange, also termed data import/export, can be obtained as: (1) direct transfer from system A to system B; (2) using a 'switchyard' mechanism, which translates incoming data structures into an intermediate structure and then from that to an

TABLE 2.6. MAIN CAPABILITIES OF SELECTED PC-BASED GEOGRAPHICAL INFORMATION SYSTEMS (modified after Fabbri and Valenzuela, 1989)

| SYSTEM | Operating system | Data structure | Digitizing capabilities | Internal database | Topologic information | Image processing | Analytical capabilities | Command language | Menu-user interface | 3-D display | Dual monitor |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *AutoCAD | PC,-DOS MAC | Vector | x | | | | x | x | x | | |
| *EASI/PACE | PC,-DOS MAC | Raster | x | x | | x | x | x | x | x | x |
| *ERDAS | PC,-DOS | Raster | x | | | x | x | | x | x | x |
| *IDRISI | PC,-DOS | Raster | x | | | x | x | | x | x | |
| ILWIS | PC,-DOS | Vector/ Raster | x | x | x | x | x | x | x | x | x |
| *MIPS | PC,-DOS | Vector/ Raster | x | | | | x | | x | x | |
| *PAMAP | PC,-DOS | Vector/ Raster | x | | x | | x | x | x | x | x |
| *PC-ARC/ INFO | PC,-DOS PS-2,-DOS | Vector/ | x | x | x | | x | x | x | x | |
| *PMAP | PC,-DOS | Raster | x | | | | x | x | | x | |
| *SPANS (PC) | PC,-DOS PS-2,-OS/2 | Raster/ Vector/ Quadtree | x | x | x | | x | x | x | x | x |
| *STRINGS | PC,-DOS | Vector | x | | x | | x | x | x | x | x |
| *TERRITORY | PC,-DOS | Vector/ | x | | x | | x | | x | | x |
| MGT. SYS | PC,-DOS | Quadtree | | | | | | | | | |
| *TOPOLOGIC | PC,-DOS | Raster/ Vector/ Quadtree | x | | x | | x | x | | x | x |

*Extracted and modified from:
Special report GIS Technology '89: Results of the 1989 GIS World Geographic Information System Survey. Published by: GIS World, Inc., P.O. Box 8090, Fort Collins, Colorado BO526, U.S.A., July 1, 1989.
Note: Also consulted The 1991 GIS World Software Survey, GIS World Inc.

external data structure; or (3) using a widely accepted intermediate neutral exchange file format. Obviously, in going from mechanism (1) to mechanism (3), we constrain the exponential increase in computer programs required for the transfer of data structures from/to each system pair.

A general description of a spatial database in a GIS is "a collection of spatial and non-spatial information in which feature attributes and feature relations are stored and are accessible". Spatial data transfer must handle the spatial data model, the non-spatial data

and the feature relations in a satisfactory manner. Usually, it is necessary to disaggregate the original data structure (i.e., the polygons) to a 'common denominator' simple structure (i.e., line segments), perform the transfer to the new computer environment of the target system, and then rebuild a new data structure in the latter, maintaining all the relationships between features.

At present, many GISs provide ample options for data transfer of vector data, raster data and of attribute tables. For PC-based systems, like the ones mentioned in Table 2.6, the most common formats are listed in Table 2.7.

In constructing a spatial database, careful planning is required to properly evaluate the cost of data transfer to integrate information from different sources. Such cost can be considerable in view of the incomplete standardization which still exists for data formats and for their documentation. It is often necessary for a user to become familiar with the characteristics of possibly unavailable systems and of their data formats to be able to import data on formats provided by those systems. Hopefully, such complexities will be disappearing with the development of new accepted data standards and more general-purpose data exchange programs for all GISs.

TABLE 2.7. SOME WIDELY USED PHYSICAL DATA FORMATS FOR PC-BASED GISs

| VECTOR | RASTER | ATTRIBUTE TABLES |
|---|---|---|
| USGS-DLG-E | USGS-DEM for CCT | dBASE-DBF |
| AutoCAD-DXF | ERDAS-LAN | dBASE-SDF |
| Intergraph-SIF | ERDAS-GIS | LOTUS-DIF |
| Arc/Info-LIN (Generate) | Arc/Info-NAS | DELIMITED(ASCII) |
| ERDAS-DIG | Windows-bitmap-bmp | |
| | ASCII | |
| | TIFF | |

## 2.5. CONCLUDING REMARKS

The construction of a spatial database is a complex and time consuming task in data analysis, in information extraction and in representation for analysis. If the final target is the development of predictive models, it becomes critical to view a GIS as a decision-making tool where the process of developing a computational model from a conceptual one can take place. Such a process is likely to bring a GIS quite far from a production system for automated cartography. Unfortunately, it is still uncommon in many public institutions to take such aspects into consideration. For this reason, often analytical tasks are performed on production systems for automated cartography at great financial and labour costs. Hopefully, the implications in the diagram of Fig. 2.11 will become common knowledge with the progress in GIS applications in spatial data analysis and modelling.

This chapter discussed data capture and data representation. Knowledge of spatial and non-spatial data structures and of their relationships in a geographical information system is essential for the integration of information for modelling. A brief mention of interaction with the data within a relational database management system was followed by an overview of PC-based GIS and image processing systems for remotely sensed data. Finally, the transfer of spatial data between different systems and some examples of

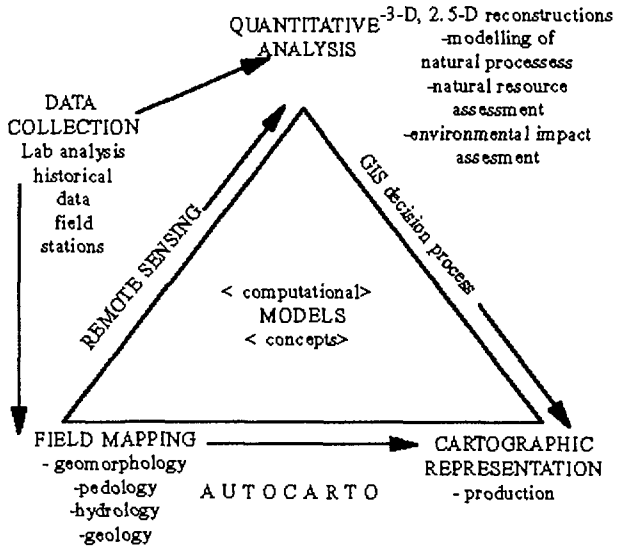



*FIG. 2.11. Representation of the dilemma between GIS modelling and cartographic production.*

powerful image algebra concluded the analysis of spatially distributed data. These concepts are fundamental for the formulation of computational spatial problems and for their solution.

## REFERENCES

Ahuja, N., 1983, On approaches to polygonal decomposition for hierarchical image decomposition. Computer Vision, Graphics and Image Processing, v. 24, pp. 200-214.

Aronoff, S., 1989, Geographic Information Systems: A Management Perspective. Ottawa, WDL Publications, 294 pp.

Brodaric, B., 1991, Map compilation with CAD for geological field mapping. Geol. Survey of Canada, unpublished manuscript (personal communication), 13 pp.

Brodaric, B., and Fyon, J. A., 1989, OGS FIELDLOG: a microcomputer-based methodology to store, process and display map-related data. Ontario Geol. Survey, Open File 5709, 73 pp. and 1 magnetic diskette.

Burrough, P.A., 1986, Principles of Geographic Information Systems for Land Resources Assessment. Oxford, Clarendon Press, 194 pp.

Chung, C.F. and Fabbri, A.G., 1993, Representation of geoscience data for information integration. Jour. of Non-renewable Resources, v. 2, n. 2.

CNR-SGN, 1991, Atti del Primo Convegno sulla Cartografia Nazionale, 2 Maggio 1991, 212+133 p. (Proc. of the 1st Meeting on National Cartography, Rome, Italy, May 2, 1991, in Italian).

Drummond, J., 1987, A framework for handling error in geographical data manipulation. ITC Journal, 1987-1, pp. 73–82.

Duda, R. O., 1980, The Prospector System for Mineral Exploration. Stanford Research Institute International, SRI, Final Report April 1980, 120 pp.

Duda, R.O., Hart, P.E., Barrett, P., Gashnig, J.G., Konolige, K., Reboh, R. and Slocum, J., 1978, Development of the Prospector Consultation System for Mineral Exploration. Stanford Research Institute International, SRI, Final Report October 1978, 193 pp.

Egenhofer, M.J., 1992, Why not SQL?. Intl. Jour. of Geogr. Inform. Systems, v. 6, n. 2, pp. 71-85.

Fabbri, A.G., Schetselaar, E.M., Crescenzi, B., and Patera, A., 1992, Some field data capture techniques for geologic mapping, production cartography and GIS processing. Proc. of European Science Foundation, ESF, Workshop on Space Time Modelling in Naturally Bounded Domains: Tools for 3-D Representation. Unpublished manuscript, 4 pp.

Fabbri, A.G. and Valenzuela, C.R., 1989, The new tools for processing geoscience information: an overview of different PC-based approaches. Proc. GIAST - 1o Workshop "Informatica e Scienze della Terra", Sarnano, Italy, October 1989, pp. 18-1-18-16.

Gorte, B., Liem, R., and Wind, J., 1988, The ILWIS software kernel. ITC Journal, v. 1988-1, pp. 15–22.

Green, A.A. and Craig, M., 1984, Integrated analysis of image data for mineral exploration. Proc. Int. Symp. on Remote Sensing of Environment, Third Thematic Conf., Remote Sensing for Exploration Geology, Colorado Springs, Colorado, April 16–19, 1984, pp. 131–137.

Griffiths, J.C., 1974, Quantification and the future of geoscience. Syracuse Univ. Geology Contr. 2, pp. 83–101.

GSC, 1991, Computerized Field Geology Meeting: Jan. 24-25, 1991. Geol. Survey of Canada, Ottawa, unpublished report, 31 pp.

Gutpill, S.C., 1991, Spatial data exchange and standardization. In, Maguire D.J., Goodchild, M.F., and Rhind, D.W., Eds, Geographic Information Systems: Principles and Applications. Harlow, England, Longman Scientific and Technical, pp. 515–530.

Healey, R.G., 1991, Database management systems. In, Maguire D.J., Goodchild, M.F., and Rhind, D.W., Eds, Geographic Information Systems: Principles and Applications. Harlow, England, Longman Scientific and Technical, pp. 251–267.

ITC, 1992, ILWIS 1.3, User's Manual, Volumes 1 and 2. International Institute of Aerospace Survey and Earth Sciences (ITC), Enschede, Netherlands, Third Edition, May 1992.

Jackson, M.J. and Woodsford, P.A., 1991, GIS data capture hardware and software. In, Maguire D.J., Goodchild, M.F., and Rhind, D.W., Eds, Geographic Information Systems: Principles and Applications. Harlow, England, Longman Scientific and Technical, pp. 239–249.

Kaufman, A., 1975, An Introduction to the Theory of Fuzzy Subsets. Vol. 1, Academic Press, New York, 409 pp.

Korth, H. and Silbershatz, A., 1986, Database System Concept. New York, McGraw Hill, 694 pp.

Mark, D.M. and Lauzon, J.P., 1985, Approaches for quadtree-based geographical information systems at continental or global scales. Procs. of Autocarto 7. Digital Representations of Spatial Knowledge, Washington D.C., pp. 555-365.

Meijerink, A.J.M., 1988, Data acquisition and data capture through terrain mapping units. ITC Journal, v. 1988-1, pp. 23-44.

Peuker, T.K. and Chrisman, N., 1975, Cartographic data structures. The American Cartographer, v. 2, n. 1, pp. 55-69.

Peuquet, D.J., 1984, A conceptual framework and comparison of spatial data models. Cartographica, v. 18, n. 3, pp. 21-33.

Poiker, T.K. and Griswold, L.A., 1985, A step towards displays of digital elevation models. Proc. Autocarto 7, Digital Representations of Spatial Knowledge. Washington D.C., pp. 408-415.

Robinove, C.J., 1989, Principles of logic and the use of digital geographic information systems. In, W.J. Ripple, ed., Fundamentals of Geographic Information Systems: a Compendium. ASPRS-ACSM, Bethesda, MD, USA, pp. 61-79.

Roscoe, S.M., 1984, Assessment of mineral resource potential in the Bathurst Inlet area, NTS 76J, K, N, O including the proposed Bathurst Inlet National Park. Geol. Surv. Canada, Open File 788, 75 pp.

Schetselaar, E. M., Woldai, T. and Fabbri, A.G., 1990, GIS training methods: a geological case study. Atti del IIo Workshop del GIAST - Informatica e Scienze della Terra. Sarnano, Italy, October 1990, pp. 202-211.

Shafer, G., 1975, A Mathematical Theory of Evidence. New Jersey, Princeton University Press, 297 pp.

Shortliffe, E.H. and Buchanan, B.G., 1975, A model of inexact reasoning in medicine. Mathematical Biosciences, v. 23, pp. 351-379.

US Department of Commerce, Bureau of the Census, 1990, Technical Description of the DIME System. In, Pequet D.J. and Marble D.F., Introductory Reading in Geographical Information Systems. London, Taylor and Francis, pp. 100-111.

Valenzuela, C.R., 1988, ILWIS overview. ITC Journal, v. 1988-1, pp. 4-14.

Varnes, D. J., 1974, The Logic of Geological Maps, with Reference to their Interpretation and Use for Engineering Purposes. U.S. Geol. Survey, Prof. Paper 837, 48 pp.

Zadeh, L.A., 1965, Fuzzy sets. Information Control, v. 8, n. 3, pp. 338-353.

# 3. VISUALIZATION

## 3.1. INTRODUCTION

The development of inexpensive high-resolution colour graphics for computers has revolutionised spatial data interpretation, exploiting the human ability to recognize spatial relationships on images. In this chapter, the hardware and software concepts used to display colour images on computer screens and on hardcopy are briefly reviewed. To begin, a typical sequence of display operations of data held in a GIS are described to illustrate how a digital picture is composed from the spatial database.

A typical cartographic image is built up with a series of display steps, as illustrated by the example in Fig. 3.1. Normally, the display is in colour, but in this case we have made a black and white picture. A desktop mapping package was used to generate this display, with data from a database built with a GIS. The database consisted of raster images, maps digitized in vector format, and point datasets with associated attributes. The data are from Eastern Shore Nova Scotia and have been used in a mineral potential mapping project. For this figure, the goal is to see where the Devonian granite outcrops lie with respect to the Bouguer gravity field, and then to superimpose the locations of known gold occurrences. The display is a fairly typical one in that several different data types and data structures are involved (areas, line and points), but the end product is a graphical rendition, initially on a monitor, and subsequently in hard copy. We will briefly run through the display operations:

(1) The Bouguer gravity values were measured at gravity stations. These are sample points on a spatially continuous gravity surface. The gravity values were first interpolated on to a grid to make a raster image. The range of continuous values was

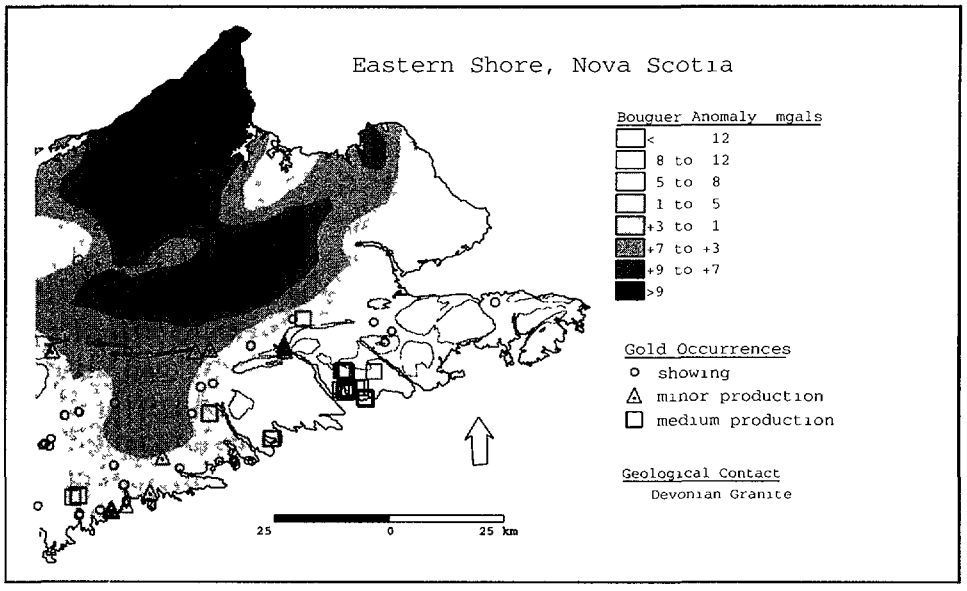



*FIG. 3.1. Layers of information combined on a digital image. The effectiveness is somewhat lost without the use of colour. An interpolated raster image (gravity), a digitized geological map (vector), and a point file with attributes (gold occurrences) are combined with a variety of graphical symbols. Annotations (legends, scale, labels, graticule) are important cartographic elements.*

subdivided into eight intervals, to simplify the picture. Grey-tone patterns were assigned to gravity classes, and the interpolated surface displayed on the computer monitor. The legend was sized and placed in position, interactively.

(2) The geological map started out in a vector format, having been digitized from a paper map. Each line was associated with attributes identifying the class of the polygon to right and left. This permitted the selection of particular lines, in this case the granite contacts, from a file containing all the geological boundaries. They were displayed using dotted lines, out of a variety of possible line symbols line thicknesses. The legend for lines was placed interactively. In addition, the vectors for the coastline were plotted as a solid line.

(3) The gold occurrences were classified into three groups, based on the amount of known gold production. The production figures were one of the fields in an attribute table linked to the points. In this case, the occurrences are discrete point objects, not sample points as in the case of the gravity stations. Three different symbols were used for display, and the legend for the points placed interactively.

(4) A scale bar, a North arrow and a title were added. Other annotations could have included a graticule, labels and textual notes.

The whole set of operations was controlled interactively with an on-screen cursor and the tools of a graphical user interface (GUI). Having composed the cartographic image on the monitor, the same graphical instructions were used to generate a PostScript file, an ASCII file accepted for printing on a variety of devices, both in black and white and in colour.

The resolution of the image on the screen is the distance on the ground corresponding to a pixel. This is often unrelated to the resolution of the data in the database. For example, if a data layer in raster format with 4000 rows and 4000 columns is stored on the disk is displayed on a screen with 1000 by 1000 rows and columns, the display is only using every 4th pixel in each direction and the original data is **decimated** to fit on the screen. Conversely for a raster on the disk that is smaller than the screen raster, some pixels are **replicated** to achieve a particular scale. When vectors are displayed on a raster device, the lines are automatically converted to raster during the display process.

The scale of the display of a cartographic image can be altered by either a **hardware zoom** or a **software zoom**. In a hardware zoom, the effect is simply that of magnifying a portion of the screen, without changing the image resolution. Individual image pixels are replicated, and as the zoom factor increases, the image becomes progressively more 'blocky'. A software zoom is often more useful but slower, because the full resolution present in the database is preserved (Fig. 3.2). A hardware zoom of a raster image that has been coverted to a vector format, for example, produces a 'staircase' appearance, whereas in the equivalent software zoom, the lines are as smooth as the scale of original digitizing will allow. Graphical metafiles, such as PostScript files, preserve the spatial resolution of the database and are particularly desirable for hardcopy display, being independent of the resolution of the graphics board and monitor.

Graphical output from GISs can be directed to devices other than video monitors, such as inkjet or laser plotters. In some cases, the contents of the video monitor (actually the display memory, as discussed below) are simply 'dumped' as a raster image whose resolution is the same as the screen raster. Alternatively, a digital plot file is created for a

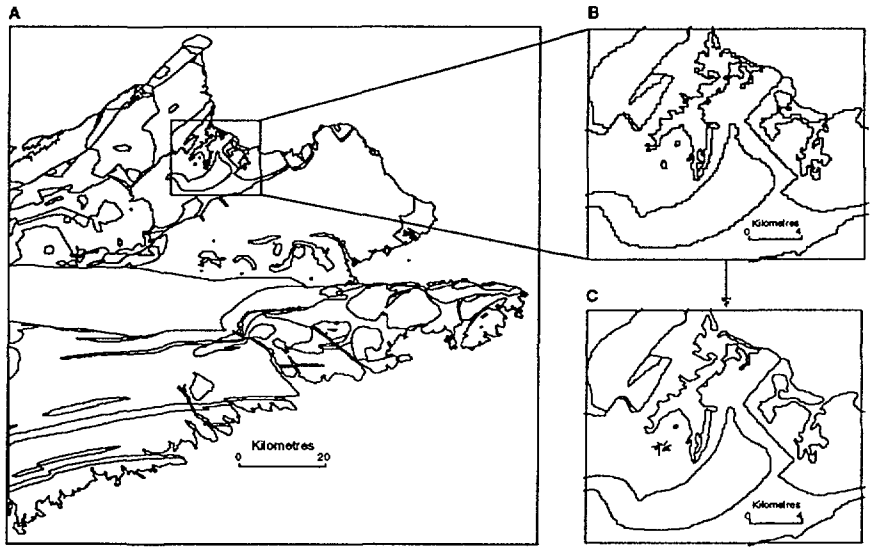



*FIG. 3.2. The distinction between a hardware zoom (B) and a software zoom (C).*

specific hardware device, such as a large-format raster or vector plotter. In other cases, a device-independent graphical file is created, often called a graphical metafile. Metafiles contain graphical instructions for raster and vector data, and can handle scalable fonts and special characters of various types. Computer programs that convert graphical metafiles to a format suitable for printing on a particular output device are known as device drivers, and are often supplied when a particular printer is purchased. A popular device-independent graphical format is PostScript, now widely used for both black and white, and colour, printers.

## 3.2. DISPLAY HARDWARE FOR DIGITAL IMAGES

The typical setup for video-display hardware is shown diagrammatically in Fig. 3.3. When displaying a data item, the display software module fills the appropriate locations in the display memory. This memory is subdivided into a number of bit planes, each one capable of storing a binary image with a fixed number of rows and columns. The number of bit planes determines the number of possible colours that can be displayed at any one time. For example, if there are 8 bit planes, then the number of colours is $2^8$, or 256. This display memory occurs physically on many machines as a single graphics card. Also on the card is a hardware lookup table (LUT) that converts the colour values on the

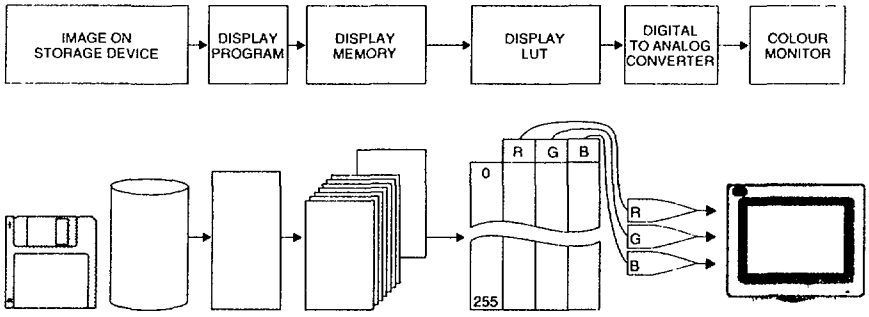



*FIG. 3.3. A typical hardware setup for graphical display.*

image into red green and blue (R,G,B) components. The digital R,G and B signals pass through a digital-to-analogue converter and finally to the display monitor. The value of each image pixel is converted to a voltage level, which governs the intensity of an electron beam at a specific location in the CRT. The CRT phosphors which occur as a lattice of individual dots, are excited to differing degrees by the intensity of the beam. On colour monitors, each pixel location has a trio of red, green and blue phosphors, and an associated trio of electron guns which separately excite the R,G and B phosphors to create colour images.

The amount and configuration of display memory on a particular graphics card varies considerably. In specifying the configuration, usually three number are used, such as $1024 \times 768 \times 8$. This means that there are 1024 rows and 768 columns of pixels, and each pixel can have a colour value from 1 to $2^8$. A configuration of $512 \times 512 \times 24$ indicates an image with 512 rows by 512 columns and $2^{24}$ colour values at each pixel. Many image processing systems use 24-bit display memory, allowing three 8-bit images to be displayed simultaneously. For example, in order to display a Landsat image, a magnetics image and a gravity image simultaneously, 8 bits (256 colour numbers) are used for each image. The Landsat values might be assigned to the red colour guns, magnetics to green and gravity to blue, or some alternative combination, producing a **colour composite** image.

### 3.2.1. Colour

The human eye has an exceptional ability to perceive subtle colour differences and to recognize colour patterns. The photoreceptor cells on the retina of the eye are of two types, the rods and cones. The cones are believed to be associated with colour perception. According to the tri-stimulus theory of colour vision, the cones are of three kinds, each one responsive to one of the three primary colours of light, namely red, green and blue. Any colour can be produced by adding together red, green and blue light in various combinations. The theory is that humans perceive colour by the relative intensity of the stimuli on the red, blue and green cones. Colour TV and video monitors work on the same principle. At each dot on the screen, a red, green and blue phosphor is present that can be excited in varying proportions. When only the red phosphor is excited, the dot appears red. When all three phosphors are excited together to the same intensity, the dot appears grey. The screen is composed of a lattice of dots, which together create a visual image in colour.

The primary colours can be arranged at three corners of a colour cube (Fig. 3.4). In the lower corner of the cube, all three primary colours are of zero intensity, and the result is the colour black. The colour corresponding to any point in the cube can be described by the displacements on the red, green and blue (R,G,B) axes. The diagonal axis cutting across the cube is an intensity axis, ranging from black at (R,G,B) = (0,0,0) at the origin, to white at (R,G,B) = (100,100,100), where intensities are expressed in percent. Elsewhere in the cube, colours are defined by their R,G,B values; (50,50,50) is dark grey, (0,0,100) is bright blue, and so on. Red, green and blue colours are called additive, because new colours are obtained by their adding them to black. The additive colours are used for video display devices, but subtractive colours are used in the dyes and inks used for many colour hardcopy devices.

The subtractive primary colours can also be represented in the colour cube. They are cyan, magenta and yellow (C,M,Y) as shown in the remaining corners of Fig. 3.4. Cyan

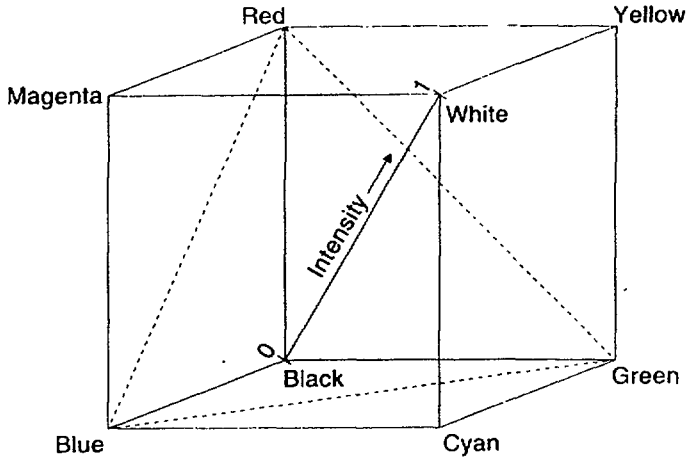



*FIG. 3.4. Colour cube model, with additive primaries (R,G and B) in one set of corners, and the complementary subtractive primaries (C, M and Y) in the other corners.*

is formed by adding blue and green light, magenta by adding blue and red, and yellow by adding green and red. Dyes and printers ink use the subtractive primaries to produce other colours by subtracting their complementary colours from white. Colours are thus defined on an R,G,B scale for video displays, and on a C,M,Y scale for most printing and plotting purposes.

One of the problems with the colour cube model is that linear changes of position within the colour cube do not lead to a corresponding linear change in colour perception by the human eye. An alternative colour model that attempts to overcome this problem is the intensity, hue, saturation (IHS, or HSI) formulation. The IHS model is geometrically related to the RGB colour cube as shown in Fig. 3.5. The intensity axis is the same for both models, increasing from 0% (black) at the origin to 100% (white) at the top. The location of a point on any plane normal to the intensity axis is defined using polar coordinates, where hue is the angular distance and saturation is the distance from the centre. Hue thus ranges in value from 0 to 360 degrees; the origin is arbitrary, but is normally either at pure red or pure blue, increasing counter-clockwise. The IHS space can be pictured as a hexcone, with pure colours lying round the top edge. Intensity is a measure of the brightness of the colour. The addition of white light produces less saturated, paler colours.

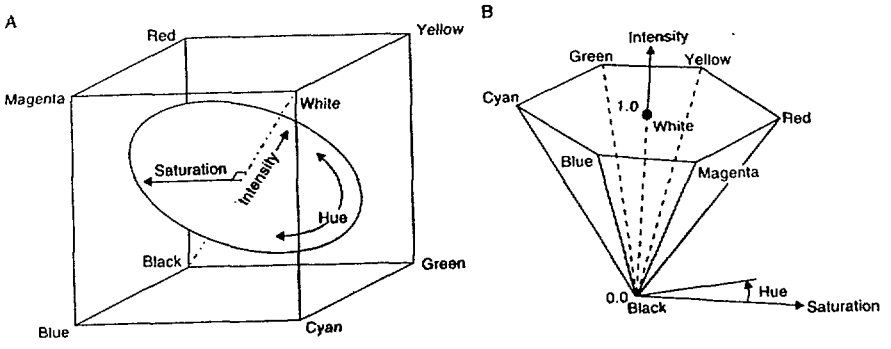



*FIG. 3.5. The relationship between the RGB colour cube and the IHS hexcone, after Harrison and Jupp (1990).*

The main advantage of the IHS model is that hue describes the human perception of colour better than a red, green and blue combination. For example, an orange hue can be modified, making it brighter by changing intensity, or paler by changing the saturation, adjustments that are difficult to guess at intuitively in R,G and B space. Similarly, saturation and intensity can be held constant, and hue can be altered to a neighbouring value in the IHS hexcone, often counter-intuitive in RGB space. In image processing systems, the IHS model is often used for colour enhancements, followed by transformation to RGB for display. It can also be effectively used for displaying multiple datasets. For example, Harris et al. (1990) produced very effective combinations of geological maps with radar. In which the geology was assigned to hue, radar to intensity and saturation to a constant value. In Fig. 3.6[1] (see Plate 1) the IHS colour model has been used to combine airborne radiometric data with airborne radar. In this case the display transformations have been carried out in two steps. First the eU, eTh and K variables were assigned to R G and B colour channels. An RGB to IHS transform was then carried out, from which the Hue was retained, but the Intensity was set to the radar reflectance and the Saturation set to a constant of 155 (on an 8-bit scale). Finally the IHS was transformed back to RGB for display. The hardcopy was produced in this case on an optical filmwriter. The image shows the Coldwell alkaline complex, a large alkaline intrusion, on the north shore of Lake Superior, as discussed by Graham and Ford (1991). The yellow circular feature is the intrusion. The town of Marathon is near the lake shore in the SE corner of the complex. Note that the radar brings out the topographic relief and drainage features, whereas the colour (hue) indicates variation in the radioelement composition of the rocks in the area. By setting the saturation to 155, the colours are pastel shades, whereas if the saturation was set to 255, the colours would be stronger. The overall result is to produce an effective combination that reveals both structural and compositional trends that aid in geological interpretation. Having briefly touched on IHS colour space, we now return to the basis of RGB space and its digital implementation.

The display hardware uses a LUT to define the actual colours that are stored in the display memory. The simplest kind of colour LUT or 'palette' is shown in Table 3.1. Each digital number is linked to a trio of colour intensities, defining the red, green and

TABLE 3.1. COLOUR LOOKUP TABLE LINKING COLOUR NUMBERS TO RED, GREEN AND BLUE INTENSITY VALUES
(the full table would contain entries for all the digital numbers from 0 to 255)

| Digital number | Red intensity | Green intensity | Blue intensity | Resulting colour |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | Black |
| 1 | 255 | 0 | 0 | Red |
| 2 | 200 | 200 | 0 | Yellow |
| . | . | . | . | . |
| 49 | 240 | 100 | 80 | Brown |
| . | . | . | . | . |
| . | . | . | . | . |
| 254 | 0 | 0 | 255 | Blue |
| 255 | 255 | 255 | 255 | White |

[1]All colour plates can be found at the back of this document.

blue. The appearance of the image can be changed instantaneously in many systems, by changing the display LUT. In image display systems with 24 bits per pixel, capable of simultaneously displaying three 8-bit images as a colour composite image, there are three LUTs, one for varying the intensity range of each of the three primary colours separately, as illustrated in Table 3.2.

TABLE 3.2. THREE LOOKUP TABLES FOR DISPLAYING THREE SEPARATE IMAGES AS A RED, GREEN, BLUE COLOUR COMPOSITE

| IMAGE 1 | | IMAGE 2 | | IMAGE 3 | |
|---|---|---|---|---|---|
| Digital number | Red intensity | Digital number | Green intensity | Digital number | Blue intensity |
| 0 | 0 | 0 | 255 | 0 | 160 |
| 1 | 26 | 1 | 220 | 1 | 170 |
| 2 | 30 | 2 | 215 | 2 | 185. |
| . | . | . | . | . | . |
| 62 | 119 | 62 | 180 | 62 | 147 |
| . | . | . | . | . | . |
| . | . | . | . | . | . |
| 254 | 227 | 254 | 15 | 254 | 38 |
| 255 | 255 | 255 | 10 | 255 | 22 |

## 3.3. COLOUR HARDCOPY PRINCIPLES

Raster plotters that use inkjet, thermal wax and electrostatic principles are now widely used for hardcopy. Each pixel in the raster image becomes one dot, or a matrix of dots, on the output medium. Optical filmwriters that transfer images on to photographic film are popular also for production of high quality hardcopy. Although hardcopy technology changes rapidly with time, the general principles of making hardcopy of raster images are likely to remain the same. The most important principle for images reproduced with inks or dyes is that colours are determined using mixtures of the subtractive primaries, cyan, magenta and yellow. Because combinations of these are limited in colour range, the appearance of mixing is simulated by a matrix of closely spaced dots made up of binary mixtures of the subtractive primaries. The eye then integrates and blurs the dots simulating particular colours.

For example, inkjet plotters create a raster of dots by squirting fine drops of cyan, magenta and yellow ink on to paper or transparent material. Whereas on a video monitor the additive primaries are added to black to produce a particular colour, on a hardcopy the subtractive primaries are 'removed' from white. At each inkjet pixel, the inkjets are either ON or OFF, giving rise to eight possible binary combinations, as shown in Table 3.3. In practice, most plotters also use a separate black ink, because the black produced by mixing C, M and Y has a brown fringe, caused by the edges of the dots not being in perfect register.

As the dot size and ink colour cannot be changed, the range of output colours produced by direct mixing is very restricted. One method used for producing a larger range of colours is a process known as **dithering**. Dithering works by using a matrix of dots to represent an image pixel. The dots in this dither matrix can be filled by any of the colours shown Table 3.3. The eye blurs the dots together, integrating the various colours to give the appearance of new colours. Dithering matrices are usually $2 \times 2$, $3 \times 3$ or $4 \times 4$. A typical dot size for a table-top plotter is about 0.2 mm diameter. This means

TABLE 3.3. COMBINATIONS OF BINARY MIXTURES OF THE THREE SUBTRACTIVE COLOURS, CYAN, MAGENTA AND YELLOW

| CYAN | MAGENTA | YELLOW | COLOUR |
|---|---|---|---|
| OFF | OFF | OFF | WHITE |
| OFF | OFF | ON | YELLOW |
| OFF | ON | OFF | MAGENTA |
| ON | OFF | OFF | CYAN |
| ON | ON | OFF | BLUE |
| ON | OFF | ON | GREEN |
| OFF | ON | ON | RED |
| ON | ON | ON | BLACK |

that each pixel occupies 0.8 mm × 0.8 mm, so that an image 250 pixels wide requires a page that is 20 cm wide. Large images are plotted using decimation and/or plotters with small dots.

A 4 × 4 dither matrix composed of sixteen dots that are either ON or OFF produces seventeen possible combinations for a given ink colour, as shown on Fig. 3.7. The dot order shown here comes from Harrison and Jupp (1990). The ordering is very critical, because inappropriate ordering can lead to distracting geometric patterns, such as herringbone or twill. The order in the matrix has been chosen to avoid horizontal, vertical and diagonal line structures. The ordering discussed by Harrison and Jupp is used in an IP system called microBRIAN, and is such that no pair of adjacent cells contains sequential numbers, and all groups of 4-adjacent cells sum to 34 (see Fig. 3.7).

To plot an image where the colours are defined by an RGB colour palette first requires a conversion to a CMY palette. For a 4 × 4 dither matrix, the intensities range from 0 (all OFF) to 16 (all ON), so that an 8-bit intensity range of 0-255 must be converted into the range 0-16. The red is plotted on an inverse cyan scale, the green on an inverse magenta scale and blue on an inverse yellow scale. For the sake of simplicity, assume that the

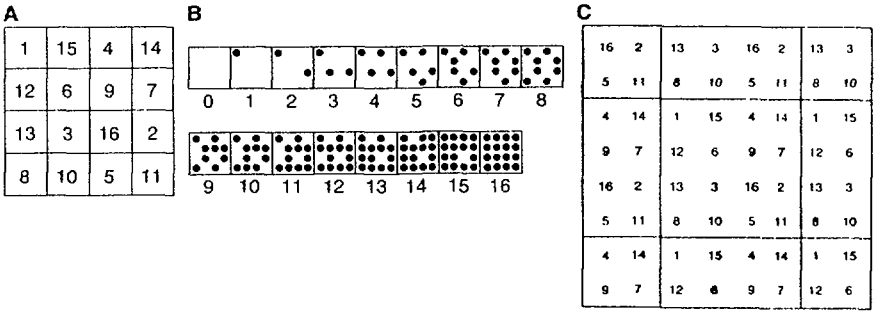



*FIG. 3.7. Arrangement of dots in (4 × 4) dither patterns used by the microBRIAN image processing system. A. Numerical code defining dot patterns. B. Dot configuration for intensity values 0–16. As the intensity increases by 1, a new dot is added, whose position is defined by the code in A. Thus the dots for intensity 5 are the same as the dots for intensity 4, with the addition of a single new dot in position 5. C. Codes for a (4 × 4) dither cell surrounded by 8 neighbouring cells (partial). Any configuration of 4 adjacent dots have numerical codes that sum to a constant value (34).*

RGB and CMY values have been compressed to the range 0-16, then an RGB of (16,0,0) is equivalent to CMY of (0,16,16), an RGB of (4,12,4) is equivalent to CMY of (12,4,12), RGB of (0,0,16) is CMY (16,16,0), and so on. Then each of the subtractive primaries is dithered with the same dither matrix, as shown in the example of Fig. 3.8. In practice, a straight conversion of RGB to CMY produces unsatisfactory results because additive colours are perceived rather differently than subtractive colours. A non-linear stretch of the image histogram improves contrast in the darker colours before plotting.

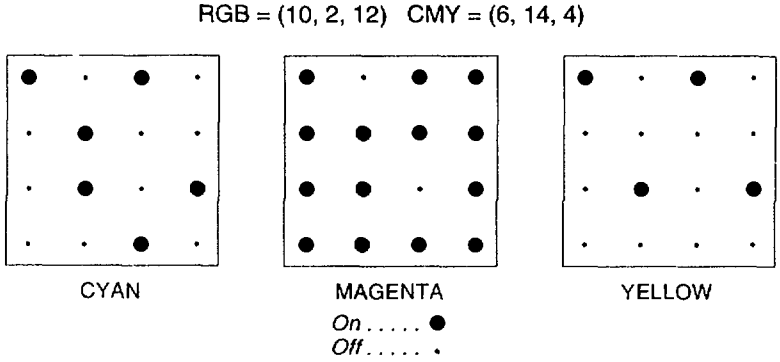



*FIG. 3.8. Dither cells for RGB to CMY for ink-jet colour hardcopy.*

## 3.4. VISUALIZATION OF SURFACES

In plan view, a continuous field variable, such as topographic elevation, is traditionally displayed as a map using contour lines. Although contours are effective for visualizing a surface, colour coded digital images, with colour palettes carefully selected to give the illusion of changing intensity, convey greater realism than contours. Images of surfaces in plan view can be made even more realistic by the use of **hill shading**, which uses information about the illumination source. Regions from where the source are not visible are in shadow. Hill-shading algorithms are found in most IP systems and GISs. Broome (1988) provides a computer program for hill-shading geophysical images, where the position of the sun can be interactively controlled with a computer mouse. Instead of shading being produced by topographic relief, the geophysical variable being mapped is treated as elevation. With hill standing magnetic dykes appear as topographic features on a magnetics image (see Fig. 3.9). As can be seen, the original image (top) shows long linear magnetic responses of the dykes. The next two images have been enhanced by hill-shading, greatly improving the definition. By changing the illumination direction, the orientation and texture of features in the image are changed and enhanced.

Even more realism can be introduced by means of perspective and isometric displays. In a perspective view, the size of objects varies inversely with distance from the viewer. Often graphic display systems create 3-D views with a 'wire-frame' model of a surface. The wire frame is a series of profiles parallel to the rows and or columns of the original grid (for a raster case), viewed by a perspective or isometric transformation. In perspective display the effect is for parallel lines to converge with increasing distance, an important depth cue for human perception. Additional realism can be added by removing the edges and surfaces that are hidden from the observer by the solid surface. Triangular meshes, as produced by Delaunay triangulation for example, can also be viewed as wire frames. Simple but effective graphical overlays are produced by 'draping' one surface

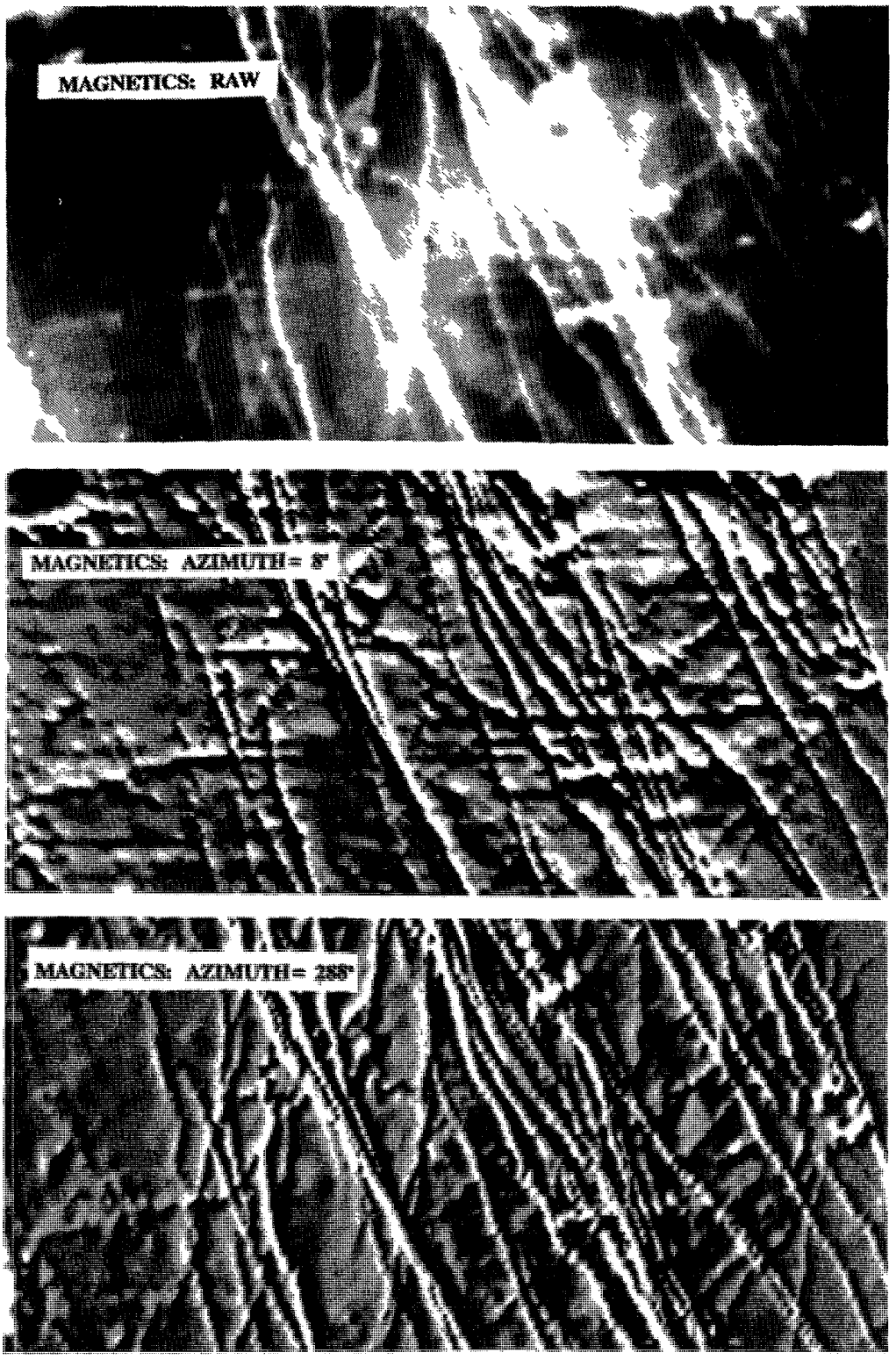



*FIG. 3.9. Shaded relief transformations for visualization of dykes on a magnetics image. A. Untransformed image. B. Shaded relief image, with illumination from NNE. C. Same, but with illumination from WNW. (By courtesy of Dr A.N. Rencz, Geological Survey of Canada.)*

over the wire frame of another surface. For example, a geological map can be draped over a topographic map, as shown in Fig. 3.10 (see Plate 2). Three dimensional representations can be rotated and viewed from different directions, yielding information about spatial inter-relationships may not be so readily perceived with other visualization methods.

Still further realism can be achieved by treating surfaces as being semi-continuous area objects, adding surface colour, texture and illumination, smoothing the geometrical artifacts of the model geometry, and creating depth cues such as the variation in haze due to atmospheric conditions, as discussed by McLaren and Kennie (1989) and Kraak (1989). These authors also discuss ray-tracing. As the name implies, this method involves following rays of light from the viewpoint into a 3-D model. Simulated rays are traced as they are reflected and refracted by objects in their path. Highly realistic views can be produced, but the method is computer intensive.

Readers interested in 3-D modelling should consult papers in the books edited by Pflug and Harbaugh (1992), Turner (1992) and Raper (1989).

## REFERENCES


Broome, J., 1988, An IBM-compatible microcomputer workstation for modelling and imaging potential field data, Computers & Geosciences, v. 14 (5), pp. 659–666.

Graham, D.F. and Ford, K.L., 1991, The integration of airborne radar and geophysical data for reconnaissance geological mapping in the Marathon-Schreiber area, Northwestern Ontario, Proc. 14th Canadian Symposium on Remote Sensing, pp. 95–98.

Harris, J.R., Murray, R. and Hirose, T., 1990, IHS transform for the integration of radar imagery with other remotely sensed data, Photogrammetric Engineering & Remote Sensing, v. 56 (12), p. 1631-1641.

Harrison, B.A. and Jupp, D.L., 1990, Introduction to Image Processing, CSIRO-Division of Water Resources, Canberra, Australia, 255 pp.

Kraak, M.J., 1989, Computer-assisted cartographical 3D imaging techniques, in Three Dimensional Applications in Geographical Information Systems, Ed. Raper, J., Taylor and Francis, London-New York, pp. 99–113.

McLaren, R.A. and Kennie, T.J., 1989, Visualization of digital terrain models: techniques and applications, in Three Dimensional Applications in Geographical Information Systems, Ed. Raper, J., Taylor and Francis, London-New York, pp. 79–98.

Pflug, R. and Harbaugh, J.W. (Eds), 1992, Computer Graphics in Geology: Three-Dimensional Computer Graphics in Modeling Geologic Structures and Simulating Geologic Processes. Lecture Notes in Earth Sciences, v.12, Springer-Verlag, Berlin, 298 pp.

Raper, J. (Ed.), 1989, Three Dimensional Applications in Geographical Information Systems, Taylor and Francis, London-New York, 189 pp.

Turner, A.K. (Ed.), 1992, Three-Dimensional Modeling with Geographic Information Systems, Kluwer Academic Publishers, Dordrecht, p. 443.

# 4. PROCESSING OF SPATIALLY DISTRIBUTED DATA

Spatially distributed data sets may consist of millions of individual observations, which must be processed correctly in order to be of use in decision making. The first goal of data processing is to make the data sets 'free' of errors, geographically linked and comparable with each other. Secondly, the aim is to enhance and extract important features and to predict new targets similar to known objects. The processing order approximately follows the steps explained in Sections 4.1–4.5. Most of these methods are applicable for both raster (pixel) and vector (point) data. Modern GIS facilities can manipulate both raster and vector data easily and can convert between these two formats.

## 4.1. PREPROCESSING

The original data in the user's database may not fulfil the criteria of 'homogenous' and 'free' of errors. Corrections and homogenization operations are therefore required.

For practical purposes, preprocessing of raw data is done in specific order, so that the correction procedure will be easier and the result will be better. The following order is recommended:

1. Description of the possible errors in (amplitudes of) the raw data due to various sources and mathematical modelling of the errors.
2. Removal of the amplitude errors through the inverse of the error function. Notice that the sample grid locations are retained at this stage. The amplitude corrections here include destriping, edge matching and levelling.
3. Geometric corrections should only be made after amplitude corrections because then there is no danger of mixing erroneous data with correct data by changing their relative geographical positions.
4. Merging, warping and mosaicking of geometrically corrected data sets.
5. Data editing: histogram operations, pan-zoom, including legends, lists, scale bars, corridors, boolean operations and data structure transformation such as vector/raster/table conversions.
6. Visualizing variable, sample and feature spaces.

Steps 1–6 are outlined in the following sections.

### 4.1.1. Corrections for data and geometry

The source of errors in the raw data sets is mostly a result of inexact calibration, instantaneous operation disturbances, inexact geographical positioning of the measuring instruments and the effects of man-made interferences.

Geoscientific data usually exist in one of four main format types: scattered field data using XY-map coordinates, airborne survey XY-profile data, aerial or orbital survey optical or scanner detector imagery in pixel array form and character data.

Field data are mostly immediately georeferenced to base maps or by GPS (global positioning system). However, old base maps may contain location errors which are difficult to correct. Generally, the location accuracy is best in the field measurements but naturally amplitude errors might occur due to drift of measuring instruments or change of analysis methods from area to area. Correction of geological, geochemical and geophysical field measurements is therefore briefly reviewed.

The way to standardize geological observations is to take samples from various formations and verify the rock type by laboratory testing.

Geochemical data sets may vary depending on sample media and analytical methods used and therefore histograms of the data from adjoining map sheets may need levelling and standardization. This is normally made by equalizing the mean/median, standard deviation and skewness of two narrow stripes on both sides of the common border line. In practice this can be done by parametric levelling, non-parametric normalization, fractile normalization and Clarke normalization (Darnley 1994).

The important geophysical corrections include Bouguer reduction for gravity measurements (Heiskanen and Moritz 1967), subtraction of temporal variation and rotation to pole for magnetic measurements (Telford et al. 1976), eliminating linear drift and thunder storm peaks from electromagnetic data (Telford et al. 1976). Radiometric measurements are frequently corrected due to variations in humidity, background, striping, pressure, temperature and height (Multala 1981). The errors in the data may cause adjacent profiles or areas look different. These differences can be minimized by the standardization methods for geochemical data mentioned above.

For airborne geophysical data the corrections are generally the same, but some additional features must be taken into account:

- Altitude, pressure and humidity variations may be high and therefore, along a profile there is a need for linear or nonlinear correction, i.e. destriping.

- The adjacent flight areas can be most conveniently levelled if the same border line is measured twice i.e. during both flying phases.

- The geographical accuracy is better if a GPS system was used for navigation.

Geometric positioning accuracy of geophysical measurements is usually so high that no additional geometric corrections are needed.

Accuracy of aerial or orbital photos and scanned images is dependent on the capacity of the sensor and conditions during imaging:

The most important characters of a sensor are:

- Radiometric resolution i.e. smallest detectable exposure change (NE $\Delta\rho$ or NE $\Delta\sigma$);
- Spectral resolution i.e. band widths of the channels ($\mu$m);
- Spatial resolution, dependent on the instantaneous field of view and flying height and velocity (m);
- Geometric accuracy, i.e. how the image geometry relates to the geometry of ground objects.

The conditions for remote sensing may vary from time to time according to:

- Illumination;
- Atmospheric absorption;
- Contrasts between the object and its background;
- Movements of the platform (carrying the sensor) in relation to the object.

These features are corrected as follows: Radiometric corrections are carried out by coefficients calculated from atmospheric models. Spatial resolution is corrected by an inverse FFT (fast Fourier transformation), if the MTF (modulation transform function) is known (see Siegal and Gillespie 1980, p. 177). However, the MTF is not often published by the manufacturer of a sensor. In this case the MTF has to be derived from the data itself.

Image striping is caused by different responses from each detector in the sensor array. For example, in Landsat MSS and TM images it appears as a repeated scan line of erroneous density. Corrections to these defects are made by various destriping methods. A quick and easy way is to match, i.e. to compare and normalize histograms of the defected lines to the histograms of the neighbouring lines. This means that the number of greytones per intervals (between maximum and minimum greytones) in a defective line is made approximately equal to the number of greytones per the same intervals in an adjacent nondefective line. Histogram matching removes erroneous scan lines effectively if the image is fairly 'smooth', i.e. without sudden peaks (see Sabins 1987, p. 240).

More sophisticated tools for destriping also exist, such as convolution and Fourier filtering. Convolution filtering provides exact knowledge of the MTF or a study of the detector response from the image. The respective error is modelled in a convolution matrix whose dimensions match the number of sensors in the imaging system. If the MTF is known, the striping will disappear when the data is multiplied by the inverse FFT of the MTF. If the MTF is not known, the amplitude of the FFT reveals as peaks those features which must be removed, usually by wedge shaped filters. The inverse FFT of the filtered spectrum is free of scan lines.

### 4.1.2. Geometric corrections

Geometric errors are a serious problem in remote sensing. *Geometric distortions* are *corrected* by modelling the distortion with *control points* and removing the error by mathematical (digital) 'rubber sheeting'.

The principles of the most common geometric distortions in Landsat MSS imagery are shown in Fig. 4.1. The flow chart for correction is shown in Fig. 4.2. GIS and image processing software usually contains tools for geometric *rectification* of RS data to any well-known map projection, i.e. Mercator, Transverse Mercator, Oblique Mercator, Space Oblique Mercator, Cassini, Lambert Conformal Conic, Albers Equal-Area Conic, Stereographic and Azimuthal Equidistant Projection (McDonnell 1979, IMSL/IDL 1992). Mathematical procedures for each of these corrections are clearly described by Bernstein et al. (1975).

The location of each data point is adjusted with respect to the corresponding point on the base map. This means that every corrected point will be georeferenced and therefore easy to combine with any other georeferenced map. However, the correction operation 'moves' the points changing interpoint distances. If the input image is a pixel image, as in the case of satellite data, the regular grid requires resampling and/or interpolation of new grid points. The most popular methods for these are nearest neighbour resampling, bilinear interpolation or cubic convolution.

### 4.1.3. Image registration

Once the first image or the data set of an area is geometrically corrected and geographically referenced it may be necessary to have two or more images in geometric conformity, i.e. in *registration* with each other. This is the usual case in mineral exploration and resource assessment, because the decision must be based on all available geodata at each point of the area. Registration is also necessary when temporal changes of the environment are studied using repetitive remote sensing actions. Extremely precise geometric corrections of the various scenes must be made so that corresponding ground objects in all scenes are assigned to the same geographical location and are thus in spatial register.

If the respective data sets (e.g. pictures) contain geographically identical locations with the same geometry of objects, these locations can be used to register a later data set with the former one. This procedure can be continued to create a large multiple georeferenced file of data sets.

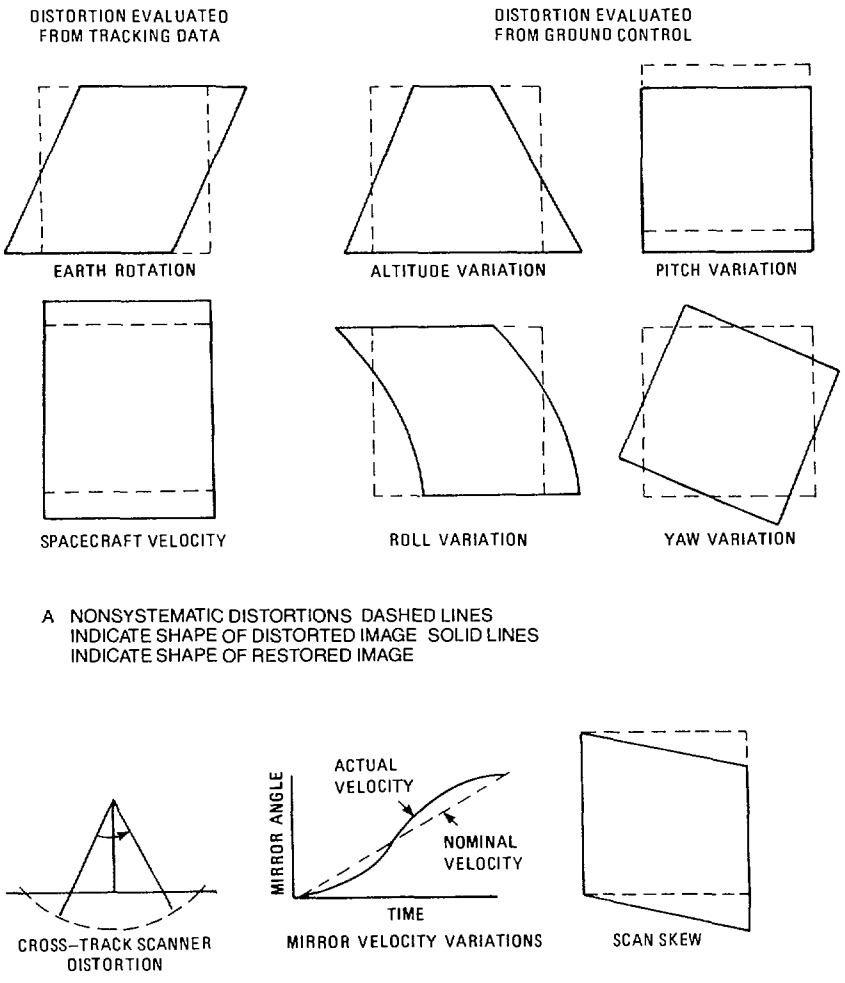



*FIG. 4.1. Geometric distortions of Landsat images. (From Bernstein and Ferneyhough 1975, Fig. 3.)*

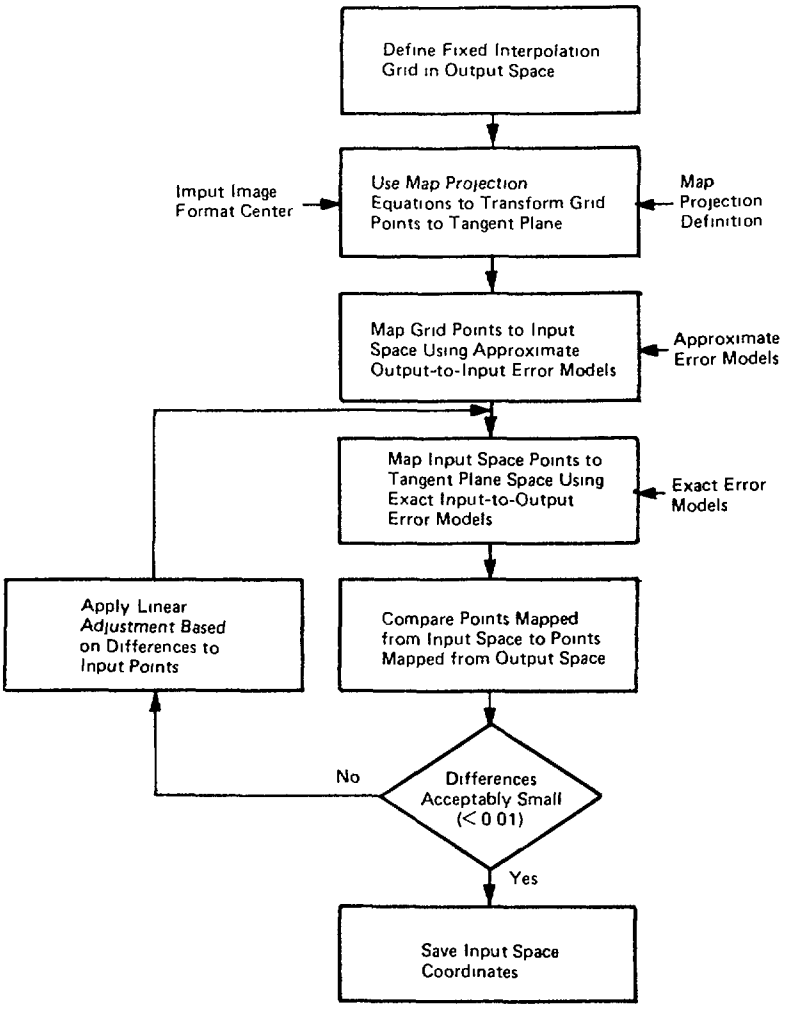



*FIG. 4.2. Flow chart for correcting geometric distortion, space-to-space mapping (Colwell et al. 1983, Figs 21–23).*

### 4.1.4. Merging, warping and mosaicking

After amplitude and geometric corrections are accomplished, the georeferenced data sets can be merged 'vertically' and warped 'horizontally'. *Merging* means forming a multivariable data set for a given area. *Warping* means compiling one map from two or more pieces which have a common border line. Both these operations are quite straightforward, and therefore they are often carried out by user-made software. However, all well-known GIS and image processing systems naturally have their own procedures for merging and warping. The problem is essentially one of digital reading and writing of corrected data sets sequentially in a framework of a specific format and data structure.

A more sophisticated approach is needed when partly overlapping multitemporal corrected data sets are combined to cover a larger area. Though amplitude and geometry are corrected in each separate set, differences may still occur which in many cases have to be minimized in order to get a 'beautiful' continuous map or an image *mosaic*. These differences may be caused by factors such as seasonal variations in the ground objects or by change of the sensor system.

The differences between two adjacent images A and B can be minimized by studying the data values in the overlapping area. Histograms of a common area $A \cap B$ can be used to formulate a linear or nonlinear operator which equalizes the histogram of the common area of image A to that of B of the same area. This operator is used to transform the whole image A to be blended with image B. However this fusion may still require equalizing of first and second derivatives in the border area. Finally, an optimal cutting line is chosen to eliminate duplicate data in the overlap region.

### 4.1.5. Data editing

*Data editing* means a number of minor digital operations, which are classically made in the photolaboratory and on a digital clipboard: *zooming* and *panning* correspond to enlarging and moving a picture. *Histogram operations* are used to adjust the greytone frequency of the histogram to make the picture informative for the eye. *Boolean operations* between two or more binary pictures correspond to the logical operations 'AND', 'OR', 'NOT'. A *corridor* is a marginal zone around a line e.g. a weakness zone around a fault. *Table to vector to raster* and vice versa format *conversions* are very important and have to be carried out with care. In modern GIS tools sophisticated *legends* can also be included in the edited pictures.

### 4.1.6. Variable space, sample space and feature space

One dimensional *variable space* is an ordered set of measurements of variable X along the X-axis. A one dimensional histogram shows the number of observations per intervals along the X-axis. A histogram is a statistical discrete approximation of a mathematical probability density function.

A two dimensional variable space is generated by all ordered pairs of measurements of two variables $X_1$ and $X_2$, spaced in a rectangular $X_1$–$X_2$ coordinate system. The name for its graphic plot is a *scatter diagram* or *scatter plot* (Fig. 4.3). Scatter diagrams are widely used to describe graphical interrelations between two variables, because statistical relations such as correlation can be 'seen' from this diagram (Fig. 4.3). Variable spaces of dimensions greater than two can be created mathematically (also in computers memory), but it is difficult to visualize higher than three dimensional scatter plots. Three dimensional scatter plots can be stereodisplayed or visualized by animated movement. Two or higher dimensional histograms can be calculated from the respective variable spaces.

If we change the roles of variables and observations, i.e. observations form the coordinate axes and a variable is situated as a point (vector) in this space, then a *sample space* is formed.

The original variables or samples may not always be suitable for further analysis. New, transformed, 'better' variables called *features* are often needed and calculated by methods explained in subsequent chapters. The corresponding space, where the original variables are replaced by the features is thereafter called *feature space (FS)*.

Variable space, sample space and feature space form the basis for most statistical operations, i.e. statistical quantities are measures characterizing these spaces. Regression, variance, discriminant, principal component, and factor analyses, classification and many other computations are carried out by studying the distribution of points in these spaces using analytical geometry and linear algebra.

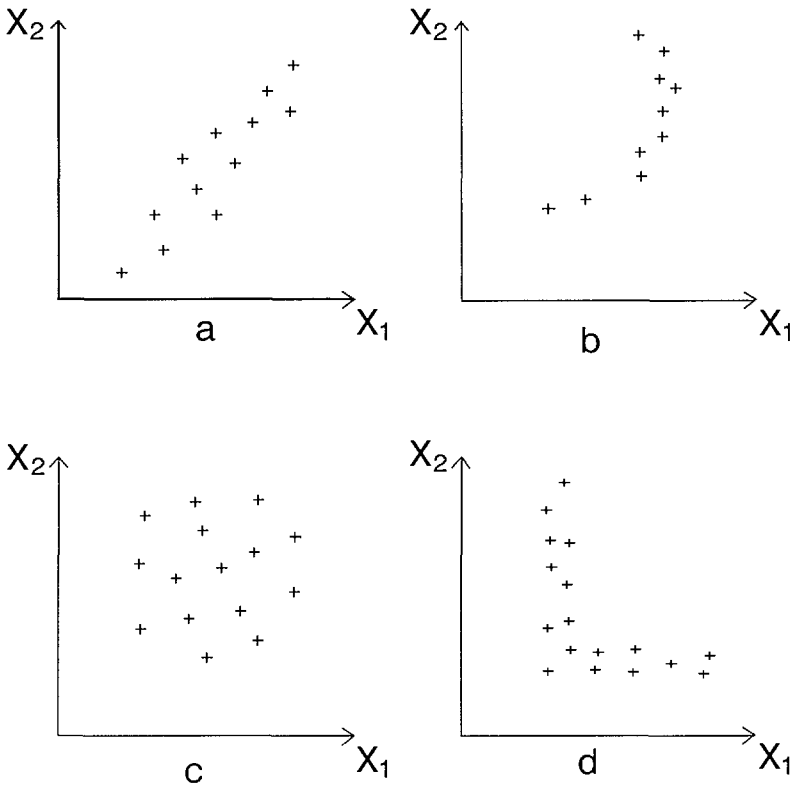



*FIG. 4.3. Scatter diagrams of observations for two variables $X_1$ and $X_2$. (a) linear correlation between $X_1$ and $X_2$, (b) nonlinear relation, (c) no correlation, (d) no correlation.*

## 4.2. SIGNAL PROCESSING

Methods originally created for the processing of sequential electromagnetic telecommunication signals are also widely used to restore, enhance and analyze 2-dimensional map and image data and to detect spatial patterns. Patterns which can be suspected as important features of targets can be enhanced by this kind of signal processing. Therefore this processing is also called *spatial feature extraction*. Spatial feature means shape and texture around one observation point or a pixel. Through spatial feature extraction the variables are transformed into new variables, features, and the variable space is therefore called feature space.

Equivalent feature extractions can be made either in the *data domain* or in *frequency domain*. Data domain means the 2-dimensional map or a picture itself where the X–Y coordinates indicate locations of the measurements. Frequency domain means an orthogonal space, where the map is approximated by a series of regular waves of different frequencies, phases and amplitudes.

### 4.2.1. Linear methods

Widely used spatial transformations used for *frequency filtering* are *Fourier* and *Hadamard* transformations, filtering in the frequency space and inverse transformation of the filtered image (Gonzalez and Wintz 1987).

A discrete 2-dimensional Fourier transform pair is given by the equations:

$$F_{(u,v)} = (1/NM) \sum_{y} \sum_{x} f_{(x,y)}^{[-j2\pi(ux/M+vy/N)]}$$

$$F_{(x,y)}^{-1} = \sum_{u} \sum_{v} F_{(u,v)}^{[j2\pi(ux/M+vy/N)]}$$

$f_{(x, y)}$ = two dimensional image
x, y = geographic location of a pixel

where u = 0, 1, 2,..., M-1, M = number of elements/line
v = 0, 1, 2,..., N-1, N = number of lines

x = 0, 1, 2,..., M-1
y = 0, 1, 2,..., N-1
j = $\sqrt{-1}$

According to Fourier theory, any map or picture data can be understood as a sum of sinusoidal waves of different frequences, phases and amplitudes. Therefore, any map can be decomposed into these separate wave components. These components are neatly expressed by a Fourier spectrum (Fig. 4.4), where the coefficients of low frequency waves lie in the central area and the coefficients of high frequencies lie further from the center. The coefficients of the wave components are also ordered according the propagation direction of the waves (Fig. 4.4).

The use of a Fourier spectrum is a way of finding periodical and directional phenomena, which can be seen as peaks or lines in the spectrum.

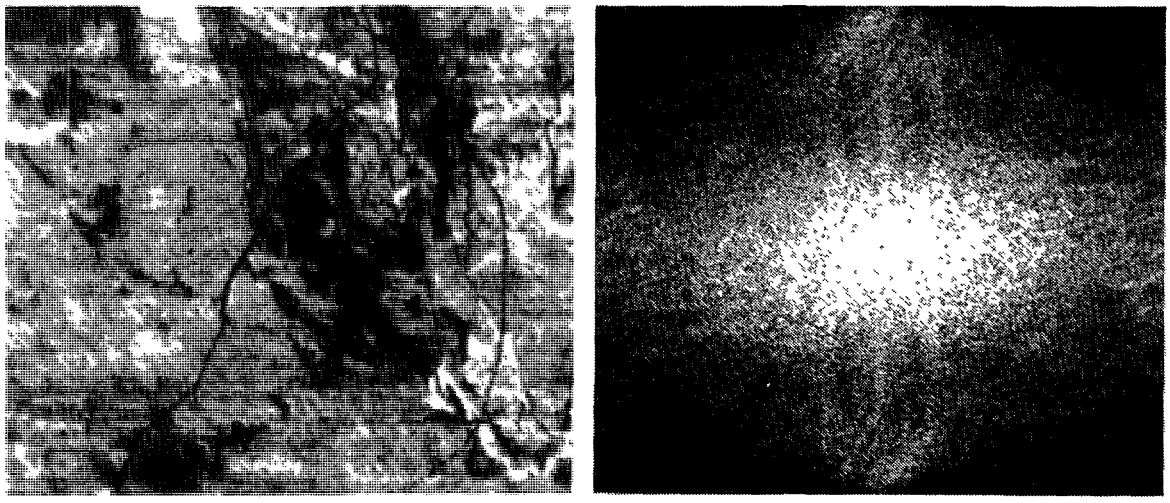


*FIG. 4.4. (a) Airborne electromagnetic in-phase map and (b) its Fourier-spectrum, amplitude.*

Fourier transforms are commonly used for decreasing some frequency components of an image. This is made through *low-pass*, *band-pass*, *high-pass* filtering (Fig. 4.5) or directional filtering. Low-pass filtering smoothens and high-pass filtering sharpens the image. Band-pass filtering enhances middle frequencies. Low-pass filtering can be used to reduce high frequency noise and high-pass filtering to sharpen a low-contrast image. Directional filtering is used for example to enhance lineaments of certain orientation. The respective low-reject, band-reject etc. filters are used if some error signals or noise of specific frequency are to be eliminated or reduced from the original image.

As was stated earlier, an image can be restored using the MTF of an imaging system. Siegal and Gillespie (1980, p. 179) show various cases of how the filter is designed when noise is absent or present.

The Fourier transform is based on sine and cosine waves, whereas a Hadamard transform decomposes an image into a sequence of rectangular waveforms. The use of the Hadamard transform is analogous to that of Fourier transform, but the Hadamard

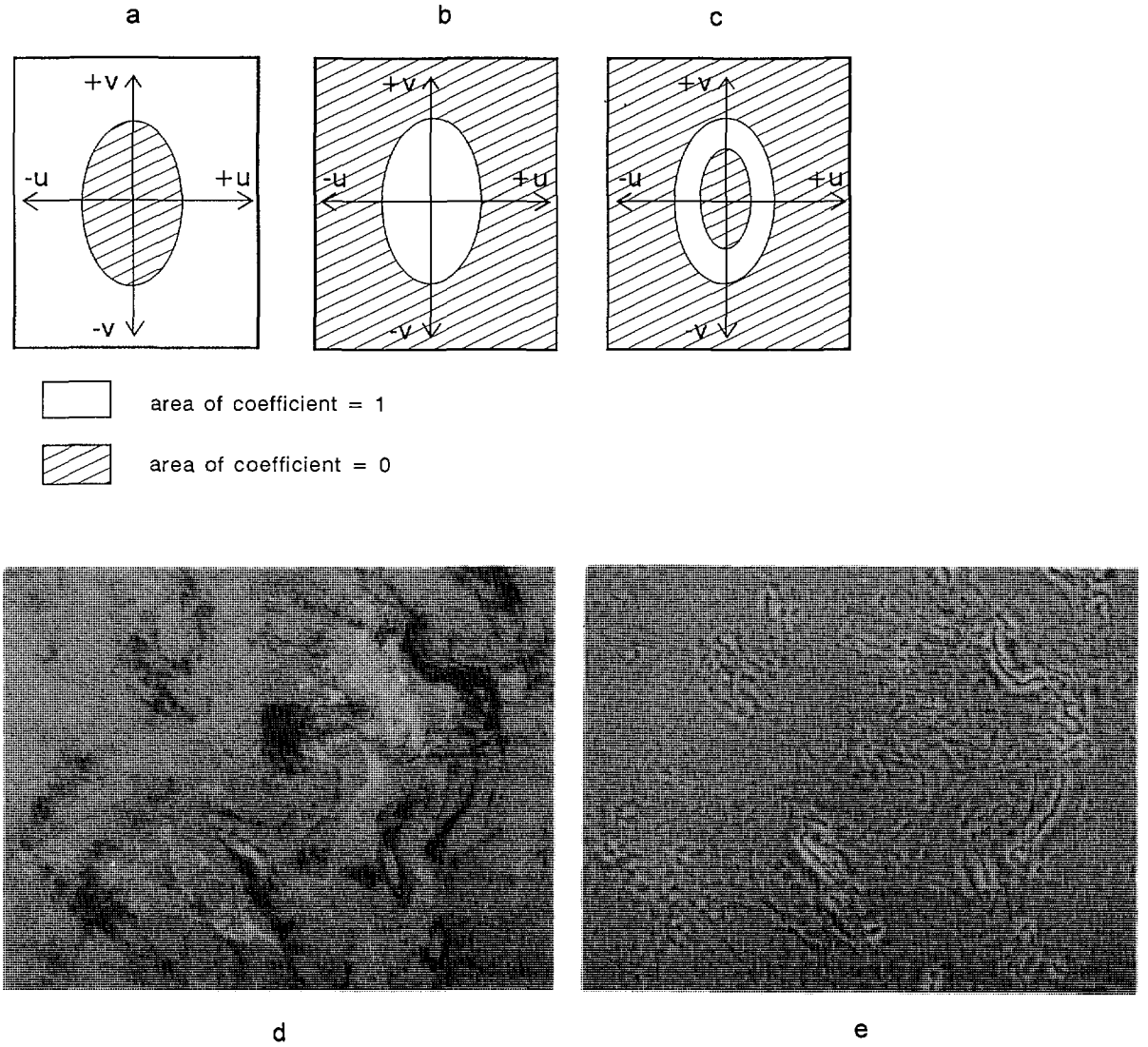



*FIG. 4.5. (a) High-pass, (b) low-pass and (c) band-pass filters in frequency space. The frequencies covered by zero coefficients are removed and the areas of unit coefficients are preserved in the resulting filtered image; (d) electromagnetic airborne in-phase component of a 60 km × 75 km area; (e) high-pass filtered version of the map in (d).*

transform works best in the case of high contrast graphic pictures or maps with fewer greytones. A small number of Hadamard coefficients can accommodate most of the variance in a dataset into the first few coefficients (Gonzales and Wintz 1987).

A discrete 2-D Hadamard transform pair of an N × N image f(x, y) is given by the following equations.

$$H(u, v) = (1/N)_y\Sigma_x\Sigma f(x, y)(-1)^{\Sigma[b\,(x)\,b\,(u)\,+\,b\,(y)b\,(v)]}$$

$$f(x, y) = (1/N)_v\Sigma_u\Sigma H(u, v)(-1)^{\Sigma[b\,(x)\,b\,(u)\,+\,b\,(y)b\,(v)]}$$

where $N = 2^n$, x, y, u and v assume values in the range 0, 1, 2,..., N-1

and $b_k(z)$ is the k:th bit in the binary representation of z. Notice that the dimension N must be equal to $2^n$ where n is an integer. For example n = 10 gives an image (of dimension) 1024 × 1024 pixels.

Fourier (and Hadamard) transformation can be used for spatial pattern recognition. Pattern P is found from an image A(P) by the following Fourier operation:

$$K(X) = F^{-1}(F(P) \cdot F^{*}(A(P)))$$

where
- K = cross correlation image
- F = Fourier transform
- $F^{-1}$ = inverse Fourier transform
- $F^{*}$ = complex conjugate of the Fourier transform

Filtering for enhancement, restoration and pattern recognition can also be made in the data domain by *convolution*. The most simple convolution is a moving average over a map. Consider moving a window with any weighting coefficients and calculating a weighted moving average over the map. This operation is a discrete convolution in the data domain. Convolution (⊗) and Fourier filtering are equivalent:

$$f\,(x, y) \otimes g\,(x, y) <=> F\,(u, v) \cdot G\,(u, v) \text{ or}$$
$$f \otimes g = F^{-1}\,(F \cdot G)$$

This short notation concerns the continuous function f which is convolved by g. The result is equivalent to multiplication of their respective Fourier spectra F and G.

### 4.2.2. Nonlinear filtering

Transform encoders can be made to adapt to local image structure. This is especially helpful if a map is composed of areas having greatly different variance or frequency contents. This is usual in geophysical field and airborne survey data. For example, continuation of a linear volcanic formation can be seen through sedimentary cover with the aid of an adaptively filtered magnetic map. Airborne, shuttle and orbital synthetic aperture radar (SAR) data also often reveals such variations because different rock types are expressed by different frequency contents of the surface morphology.

Simple *adaptive filters* may be designed by modifying the formula of the high-pass convolution filter for an image X(x, y) (Fahnestock and Schowengerdt 1983)

$$Y(x, y) = a\,[X(x, y) - (1 - b)\,X^m(x, y)] + c$$
$$= a\,[(1 - b)\,(X(x, y) - X^m(x, y) + b\,X^m(x, y)] + c$$

This is a normal high-pass filter in a K by K window at location (x, y). Y(x, y) is the output, X(x, y) is the input and $X^m(x, y)$ is the average of pixel values in the moving window. The scalar b controls the relative proportions between the input and the high-pass component. The scalars a and c control total contrast and mean of the output Y(x, y).

Peli and Lim (1982) suggested following filter which is adapted through functions $T_1$ and $T_2$ depending on the local mean.

$$Y(x, y) = a[T_1\,(X^m(x, y)) \cdot (X(x, y) - X^m(x, y)) + T_2\,(X^m(x, y))] + c$$

Proportions of high-pass ($X - X^m$) and low-pass ($X^m$) in Y are balanced by the formulation of the functions $T_1$ and $T_2$.

Harris (1977) proposed that the rate of high-pass is divided by the local standard deviation S(x, y) in the window

$$Y(x, y) = A[1/S(x, y) \cdot (X(x, y) - X^m(x, y))] + C$$

This filter increases high-pass variation in smooth areas and decreases high-pass in rough areas. Both the window size and the weighting function may be formulated to adapt to the local circumstances in the map.

### 4.2.3. Mathematical morphology

Mathematical morphology is an approach to image processing which is based on the theory of sets. In a digital image, a group of objects of a same type is represented by the set of pixels belonging to the objects. Those pixels are located within a larger set of pixels comprising the entire image space in which other types of objects and the background are also identifiable as different sets of pixels. Exploration and measurement of the desired objects is obtained by scanning the image with small images or probes (also termed structuring elements) that provide measurements of specific characteristics, such as dimension, orientation, and distribution, to enable the geometrical separation of the objects. When a model is available for the interpretation of the geometrical properties of the objects in an image, the approach can provide the background for extrapolation and for prediction. According to Serra (1982), set theory can be used to construct a picture algebra by developing criteria and models for image analysis, for example considering the morphology of binary images (black and white), of grey-tone functions, and of random sets (Matheron, 1975).

Most of the computations to be performed on images for morphological characterization are of local nature, for example, edge detection, thresholding, thinning, skeletonizing, and object counting. Thus a function is evaluated that, for each pixel, takes into account the values of a subset of neighbouring pixels (not excluding the pixel itself). Local operators or neighbourhood operators may be used for detecting both geometrical and

topological features, in order to 'understand' the image. This section is a brief introduction to morphological transformations or morphological filtering which exemplifies how simple transformations can be sequenced for feature extraction in geological applications of digital remote sensing. In addition, a brief discussion is provided of the usefulness of mathematical morphology filters in modelling data integration and in spatial analysis in general within geographical information systems.

Simple examples of shape-sensitive filters can be generated by small binary kernels which identify the pixel positions, within a neighbourhood, which are used to compute a function whose result will decide the value that an image pixel, corresponding to the centre kernel pixel, will be assigned. Figure 4.6(a)–(i) shows kernels with centre pixels underlined to simplify their mapping. The "1's" indicate pixel locations which are used in the computation, e.g., to compute the maximum (minimum) value of the image pixels corresponding to them. The "x's" correspond to the pixel values that are ignored. We can also consider the kernels in Fig. 4.6 as small images, and transform one by using either itself or another image kernel. For instance, by using the kernel in (a) to compute the maximum value for all the corresponding pixels in (b), we obtain the image in (c). A similar result is obtained transforming (g) with (h) to generate the image in (i). Sweeping an image with a structuring element, means to translate the latter so that its centre pixel overlies all the image pixels in sequence and for each overlay, a pixel value is computed in the resulting image. The computation of minima is termed an 'erosion' while the computation of maxima is termed a 'dilatation': this conveys the concept of shrinking or of expansion, respectively, of the boundaries of the objects in the image. By sequencing such transformations, i.e., eroding an image and then dilatating its erosion (or dilatating it and then eroding its dilatation) sets of transformed images are obtained which are contained in (or contain) the original or the previous image in the sequence. These transformation sequences are termed opening and closing, respectively. Their effect is of 'sieving' the objects according to their size in one or two directions of the image raster (generally square, sometimes hexagonal). Kernels (d) to (i) in Fig. 4.6, identify particular directions or senses of direction; kernels (a) to (c) are isotropic and work in several directions.

For binary images (two valued, e.g., 0 and 1, or white and black) topology sensitive filters can be used which map all the pixel neighbourhood configurations which correspond exactly to the kernel configuration. In Fig. 4.6(j)–(m), the mapping of small convexities, small concavities, vertical straight segments of length 5 pixels, and oblique segments of 5 pixels are shown. Matching 0's and 1's are computed while the x's identify "don't care pixel value positions."

Appropriately designed sequences of transformations can represent strategies for the extraction of the desired information or of the desired objects from an image. The approach leads to the initial separations of particular aspects in a 'granulometric' sequence for later reassembly or reconstruction of the objects having acceptable geometrical or distributional characteristics.

A simple example of morphological filtering of a one-dimensional signal by a 3-pixel long horizontal kernel is shown in Fig. 4.7. In the illustration it becomes evident how closing identifies the dark valleys that can be isolated by subtracting the original signal from it. Conversely, opening leads to the identification of the bright peaks which are then isolated by subtraction from the original signal. The analysis of images using structuring elements of sequentially greater lengths and of different orientation leads to complex structural analyses as it will be discussed later when reviewing some applications.

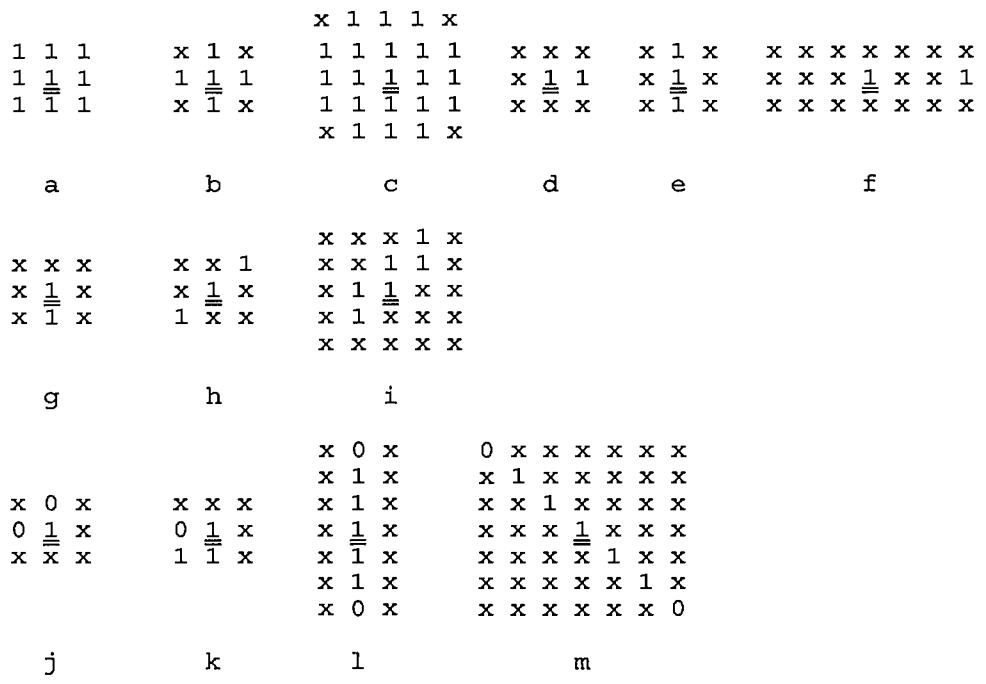



FIG. 4.6. *Examples of structuring elements kernels used in mathematical morphology filtering. (a)–(i) can be used for either binary or grey-level images (erosion, dilatation, opening, closing); (j)–(m) are topology-sensitive kernels that can be used for binary images (hit-or-miss transformations such as connectivity number, histogram of given directed lengths, etc.).*

An application of morphologic filtering, termed 'top-hat' transformation, is described in Fig. 4.8 (see Plate 3). The image of the aeromagnetic anomaly pattern, for a study area around Bathurst Inlet in the Northwest Territories of Canada, which was provided by the Geological Survey of Canada as an array of average gamma values for square cells of side 812 m, was resampled to 100 m pixel resolution. The values of the gamma readings were rescaled between the values 0 and 255, i.e. to one byte. The grey level image for a 800 pixels × 600 lines subimage is shown in Fig. 4.8(a), where the sets of bright peaks identify high readings which correspond to highly magnetic objects (diabase dykes and gabbro sills). Several structural trends can be observed, NW-SE, E-W, and some N-S orientations. The extraction of bright peaks in all directions was obtained using the circular (octagonal) structuring element in Fig. 4.6(b) twice (i.e., a circular element of diameter 9 pixels). Fig. 4.8(b)–(d) show the result (the grey level were stretched between 0 and 255) of the erosion (max) by it, of the opening (max and min), and of the subtraction of the opened image in (c) from the original image in (a). The binarization of the image of the subtraction, shown in Fig. 4.9(a) (see Plate 4), was obtained by thresholding (grey level slicing) it at values greater or equal 2. This complex three-step transformation is termed top-hat transform in mathematical morphology (Serra, 1982).

To verify the correspondence of this feature extraction process, the overlay is computed in Fig. 4.9(b) with the structural data about faults, diabase dykes and the square cells around mineral occurrence pixels digitized from a 1:250 000 geological map (Roscoe, 1984). From this illustration it can be seen that the E–W trend of dykes is underrepresented in the map. A systematic enhancement of aeromagnetic data was not used to produce the reconnaissance mapping which produced the geological map. Evidently, contour maps of aeromagnetic anomaly did not allow a satisfactory visual

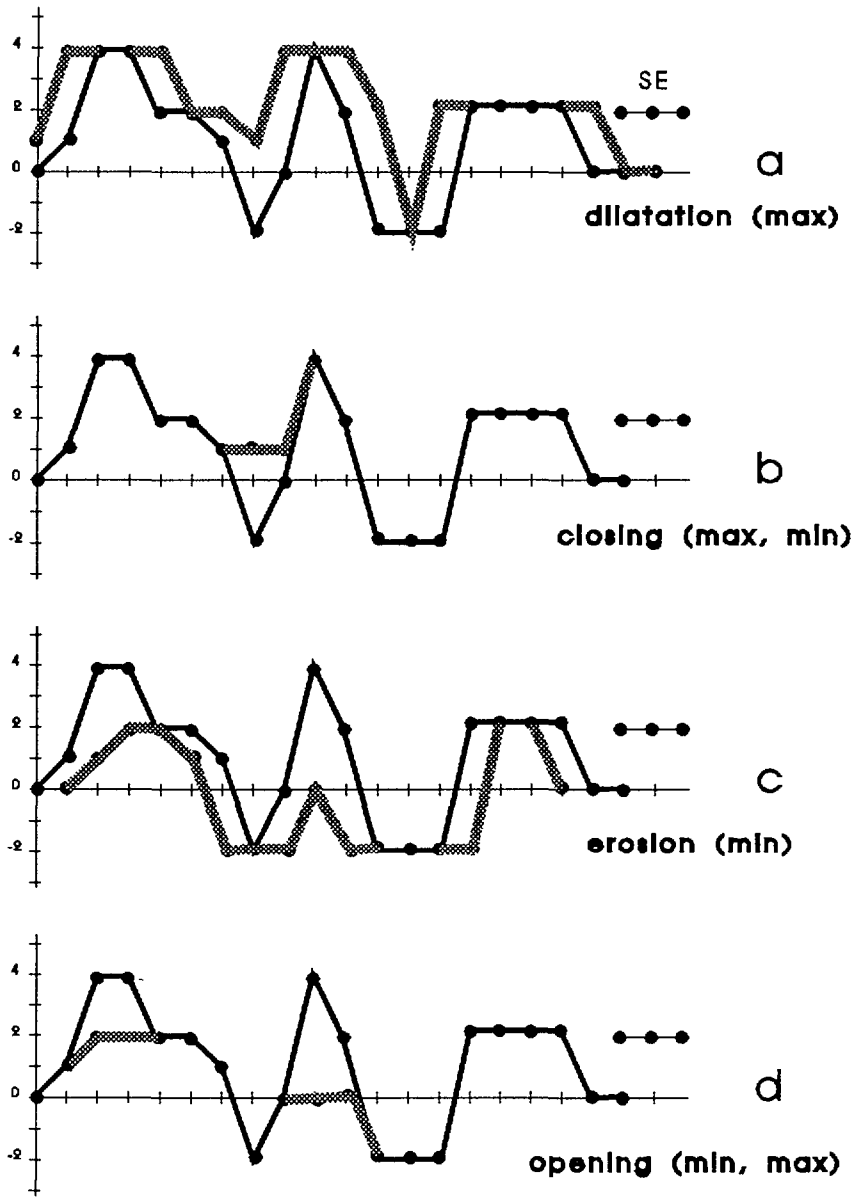



*FIG. 4.7. Examples of morphologic transformations on a one-dimensional signal (or cross-section of a grey level image) by a linear structuring element of length 3 pixels. (a) in grey the result of a dilatation (max); (b) in grey the result of a closing, i.e., an erosion following a dilatation; (c) in grey the result of an erosion; (d) in grey the result of an opening, i.e., a dilatation following an erosion.*

feature extraction. Many observations of E–W trending diabase dikes remained as field notes and were not represented on reconnaissance maps during fieldwork at 1:50 000.

The isolation of the E–W trend in the aeromagnetic anomaly image of Fig. 4.8(a) was performed in Fig. 4.9(c) by computing a top-hat transformation using a vertical structuring lineament similar to the one shown in Fig. 4.6(e), but of length 7 pixels (equivalent to iterating (e) three times). The binary image of the extracted horizontal anomaly in Fig. 4.9(c) was then 'cleaned' of all short horizontal segments of length less than 9 pixels. This is displayed in Fig. 4.9(d), overlaid with the structural data. The two Figs 4.8 and 4.9 (Plates 3 and 4) illustrate some common transformations which can be used in image analysis.

The following is a review of some of the fundamental research on the general principles applied in mathematical morphology and on the relationships between conventional and morphologic filters. Studies on the properties of morphologic filters demonstrated that translation invariant and increasing transformations can unify linear, median, and order-statistics filters under one roof (Maragos and Schafer, 1985; 1987a, b). This is done by representations as minimal combinations of morphological erosions or dilatations. Applications of morphologic filters can provide'systematic algorithms for image processing and analysis for tasks such as nonlinear image filtering, noise suppression, edge detection, region filling, skeletonization, coding, shape representation, smoothing and recognition. Several results reviewed by Maragos (1987) lead to a unified image algebra for operations and processes. Recent tutorials on both binary and grey level image analysis using mathematical morphology were provided by Maragos (1987) and by Haralick et al. (1987). The latter two articles offer extensive background on the properties of morphologic filters to be employed in the construction of processing strategies.

In remote sensing, two applications clarify how processing strategies can be constructed using morphological filters in combination with more conventional image processing techniques. Destival (1986) used mathematical morphology to extract the patterns of villages and roads from a SPOT scene. She observed the textural and tonal appearance of fields, roads and villages in the different channels of the scene under study (highly contrasted villages, homogeneous fields, roads as clear but interrupted peaks of varying width). A succession of gradient operators, skeletons and openings isolated the villages. Top-hat transformations were then used to clear first the dark noise and then to extend the bright roads. Binarization of road pixels was followed by closing filtering to connect them and to reconstruct interrupted roads into a connected network. The relative importance assigned to different pieces of roads enabled to optimize the sequence of steps to generate an acceptable village and road pattern extraction. The complexity of the analysis and the identification of several alternative strategies and results in this application revealed the duality between ground context knowledge and the experience gained of the digital context: it has to be resolved by interpreting one in terms of the other.

Flouzat and Moueddene (1986) developed a framework to aid geological interpretation of a Landsat MSS scene (bands 5 and 7) using mathematical morphology for the study of the pattern of an anticline in a forest vegetated area. Sites of strong slope with forest and shaded areas corresponding to low reflectance pixels were extracted by classification. Morphologic filtering was used to increase the compactness of pixels (and of pixel clusters) before classification (nested sequences of grey tone openings and closings, e.g, one erosion, followed by two dilatations, followed by another erosion, by a same structuring element). Binarization provided the noisy silhouettes of the anticline which was then subjected to nested opening and closing to eliminate the black and the white noise. This was followed by skeletonization (line thinning which maintains pixel connectivity), pruning (the elimination of small branches in thin line objects), and the elimination of isolated pixels to generate the thin skeletons of the silhouettes. The latter were then direction labelled (i.e., each pixel was assigned a direction value within its neighbourhood) to extract only the silhouettes with a particular orientation. The strategy lead to the identification and subsequent filling of holes corresponding to shaded areas (umbras, i.e., a disturbance caused by the orientation of relief elements) and to the determination of the main structural directions.

Durand and Flouzat (1985) used shape characterization by mathematical morphology to quantitatively analyze the visual aspect of the classes resulting from the classification of a

SPOT image over rice fields in Mali. In particular, they performed textural analysis and structural reconstruction of mixed classes to obtain more compact entities with sharp edges, therefore improving their identification. The same technique was applied by Fabbri et al. (1993) to the results of supervised and unsupervised classification of Landsat MSS in northern Canada in order to identify viable general-purpose processing strategies.

Poujade and Laurore (1990) proposed a classification method which uses the contribution of both spectral and textural criteria. An optimal morphologic operator was identified by using different sizes and shapes of structuring elements and the associated statistics. The application problem was the extraction of urban zones and specific crops from SPOT satellite images. Several textural classes extracted were compared (i.e., crossed or correlated) with the spectral classes by computing the number of pixels belonging to both and by analyzing variation of the pixel properties with the different sizes of structuring elements. This enabled the identification of maxima of correspondence.

In a more general approach to texture analysis, Dougherty et al. (1989) proposed to use local morphologic granulometries (standardized sequences of opening and closing to obtain a normalization of size distributions) to partition textural domains. Rather than constructing a single size distribution based on the entire image, local size distributions are computed over moving windows within the image. Regions of various textures can be isolated when the local statistics is found to be homogeneous over a given subregion.

Martel et al. (1989) analyzed the pattern of free air gravity anomaly for gridded data at 1/4 degree intervals over the Indian Ocean (0° to 45° S, 55° to 105° E). The data were converted into 0-255 integer values (bytes). They used morphological filters (opening and closing by differently oriented linear structuring elements) to perform: (1) peak and valley detection, (2) orientation analysis, and (3) period-icity search. They observed that, unlike the usage of fast Fourier transforms, morphologic processing using local operators is not affected by the physiography of the area studied, which is subdivided in zones by ridges and through. Also, the simplicity of the processing makes it easier for the final results to be interpreted. Fabbri and Kushigbor (1989) used a similar approach in a study of the aeromagnetic anomalies shown in Figs 4.8 and 4.9 (Plates 3 and 4).

Morphological filters have also been used for the elimination of noise from satellite imagery such as the speckle in radar images (Safa and Flouzat, 1989) and for destriping defective SPOT images (Banon and Barrera, 1989).

The variety of applications described here, is only an indication of how shape analysis can contribute to the modelling of spatially distributed data. Early examples of spatial models for data integration in mineral exploration and for microtextural analysis were discussed by Fabbri (1984) who generalized image processing to geological applications. Mathematical morphology techniques can additionally be seen as generalizations of many spatial transformations common in geographical information systems and in spatial reasoning by expert systems, as discussed by Fabbri (1991).

## 4.3. SPECTRAL ENHANCEMENT

In addition to the extraction of spatial features, such as frequency components, texture, geometrical and morphological properties from a given variable, it is sometimes necessary to create new variables by enhancing the individual observations separately from their surroundings. Spectral enhancement or *spectral feature extraction* is carried out by

operations which are made for only one sample or pixel at a time. Through the following operations the variables are transformed into features and therefore the original variable space is spectrally transformed into feature space.

### 4.3.1. Histogram transformations

A *histogram* of a map or an image is a list of numbers each indicating how many measurements fall into intervals between maximum and minimum of the data. This list is often graphed by a sequence of columns or an X–Y plot. A histogram is a discrete analogy to the probability density function.

The purpose of histogram modification is mainly to improve the visual impression of an image or to prepare data for further statistical analyses which require a normalized distribution. A visual impression of an image is 'good' if the grey level histogram is nearly a normal distribution having an 'appropriate' mean and standard deviation. The mean is set equal to the mean greytone and variation is scaled to satisfy the perception of a human eye (Fig. 4.10). Statistical treatments like factor analysis require that the histogram be (0, 1) standardized, i.e. the mean must be equal to 0 and standard deviation equal to 1.

A histogram can be modified by a *level transformation function* (Fig. 4.10). Its discrete form is called a *look-up-table*. Any form of the histogram can be obtained by designing and applying an appropriate level transformation function (Gonzales and Wintz 1987). Image processing and GIS software usually have interactive tools for histogram modification.

There are some special cases of histogram modification which are often used: histogram equalization and linearization produce a uniform histogram for the output image. Normalization produces a normally distributed output histogram and optimization transforms the histogram so that the output image is the best possible for the human eye.

### 4.3.2. Spectral enhancements for multichannel remote sensing data

Remote sensing is based mainly on elecromagnetic wavelengths between 0.3 $\mu$m – 60 cm. Most alteration minerals can be identified by wavelengths between 0.4–2.5 $\mu$m with a spectral resolution 4 nm, i.e. 0.004 $\mu$m. Therefore 525 channels are needed ideally for such mineral identification. Nearly all minerals can be identified by wavelengths 1–25 $\mu$m using a spectral resolution of 4 nm. Airborne and satellite sensors usually have less spectral resolution capacity because the channels are wide and the number of channels is limited. Therefore one has to study carefully spectral reflections from the object under consideration. For instance the Landsat TM sensor has 7 channels each covering $\geq$1000 nm.

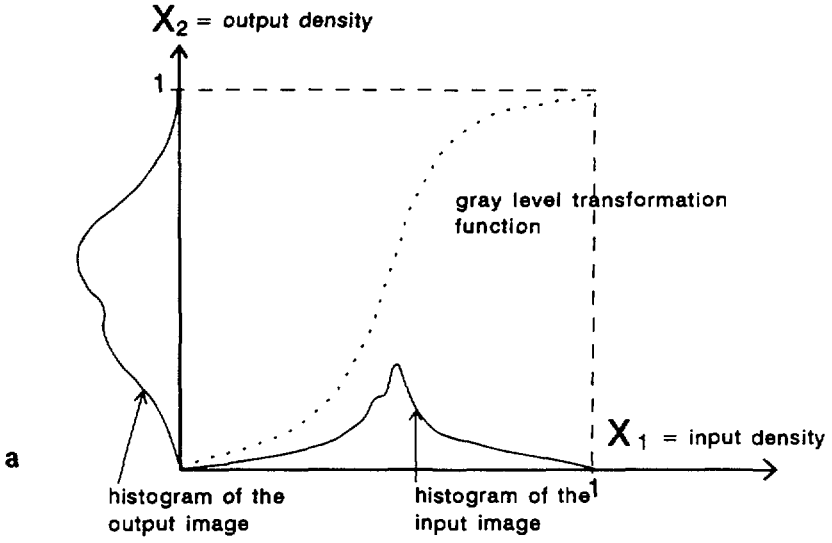
The detailed reflectance model of an object is therefore designed according to the data from imaging or field spectrometers. This model will thereafter be approximated by the 'rough' channels of the satellite or other sensor. In many cases this has produced successful results (Simpson et al. 1991, Smith 1977, Goetz et al. 1982).

The approximation of the fine spectral reflectance model by a few wide channels is also made by band ratioing. For example, the following Landsat band (Lillesand and Kiefer 1987, p. 594) ratios are used:

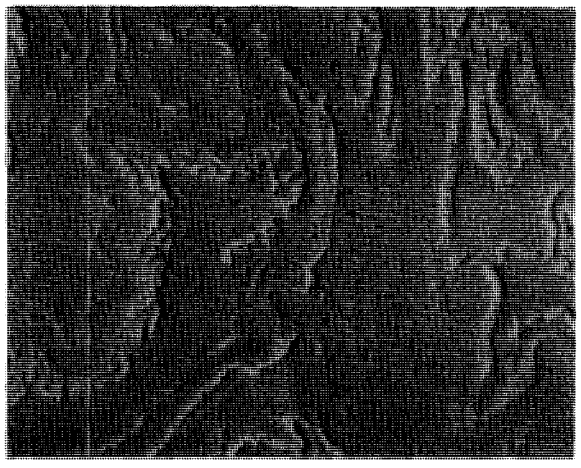
TM 7/5
7/4 for alteration mapping
5/4
3/7
1/5 for vegetation stress analysis
7/6

MSS 4/6
5/6 for vegetation mapping
4/7
5/7

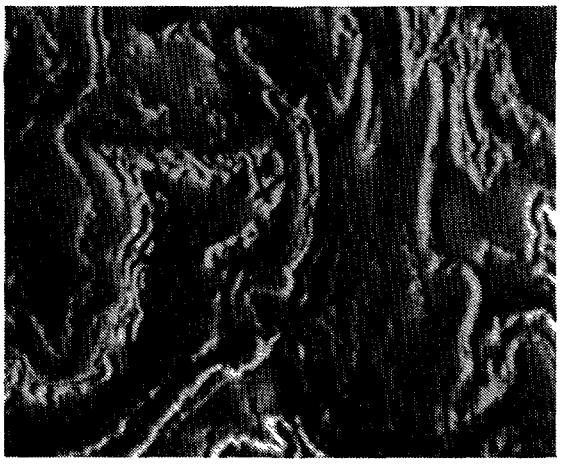
4/5
5/6 for lithologic discrimination,
4/7 h-altered/nonaltered rocks
6/7




a



b



c

*FIG. 4.10. The effect of the histogram transformation of an image. (a) Grey-level transformation, (b) input image, (c) output image.*

Ratioing of the bands reduces the effect of brightness variations due to topographic slope. At the same time, however, ratio images increase noise thus making interpretation more difficult. Band ratioing is also a widely used technique for gamma radiation measurements, because the damping effect due to water can be reduced in this way. U/Th, K/Th and U/K are the usually computed new variables (Kuosmanen et al. 1988).

Vegetation indices are vegetation-sensitive combinations of red (Red, 0.6–0.7 $\mu$m) and near infrared (NIR, 0.7–1.1 $\mu$m) channels of sensors. The red band records chlorophyll absorption and NIR-band records the strong reflectance of healthy leaves. The simple *vegetation index* (VI) and the *normalized difference vegetation index* (NDVI) are defined by Lillesand and Kiefer 1987, p. 597 as:

$$VI = NIR / Red \quad .$$

$$NDVI = (NIR - Red) / (NIR + Red)$$

These are extensively used for monitoring large areas of green vegetation (see Sabins 1987, pp. 321–323 and plate 14 A).

For the NOAA (National Oceanic and Atmospheric Administration) AVHRR (Advanced Very High Resolution Radiometer) images (Lillesand and Kiefer 1987, p. 594) these correspond to (B1 and B2 are the intensities of respective bands):

$$VI_{NOAA} = B2 / B1$$

$$NDVI_{NOAA} = (B2 - B1) / (B2 + B1)$$

For the Landsat TM imagery they are calculated as (TM4 and TM3 are the intensities of the respective bands):

$$VI_{TM} = TM4 / TM3$$

$$NDVI_{TM} = (TM4 - TM3) / (TM4 + TM3)$$

The normalized difference vegetation index, which has been successful in mapping biomass in equatorial regions, does not work at all with Boreal coniferous forests (Häme et al. 1992). Therefore Häme et al. (1992) developed a more sophisticated *Boreal Biomass Index* using Landsat TM data:

$$BIOMASS = C * V$$

$$Ln(V + 1) =$$

$$-43.1 + 2.18\ TM2 - 3.38\ TM5 + 16.7\ TM3/TM2 - 0.81\ TM4/TM3 + 24.8\ TM5/TM1 + 43.7\ TM5/TM2 - 53.7\ TM7/TM1 + 12.1\ TM7/TM3 - 0.02\ TM3*TM2 + 0.04\ TM5*TM2 + 0.01\ TM7*TM5$$

where

| | | |
|---|---|---|
| V = | | stem volume ($m^3$/ha) |
| TM1-TM7 = | | the intensities in channels 1 through 7 of the Landsat Thematic Mapper image |
| Ln = | | natural logarithm |
| C = | | constant, suggested value 90 |

### 4.3.3. Simultaneous processing of many variables/features

Modern computer facilities have remarkably increased our capacity to combine many geoscientific variables or inferred features. Geochemical analyses from soil and bedrock samples may produce data for up to 40 variables, geophysical airborne or field surveys can produce 20 variables and satellite sensors 10 variables or more particularly if multitemporal data is included. Imaging spectrometry may lead to as many as 256 variables!

Simultaneous studies of groups of these variables can lead to improved recognition and definition of 'fingerprints' for exploration targets, geological formations and environmental hazards. Simultaneous studies are carried out by sandwiching maps of different variables and by statistical studies of variable, feature or sample spaces. In normal processing strategy these are both made alternately by digital image processing. Processing strategy is discussed later in the text in Section 4.3.5.

A statistical approach to studying many variables simultaneously can be of value in:

(a) studying relations between variables/features or samples and increasing contrasts between variables;
(b) decreasing dimensionality, i.e. decreasing number variables by replacing them by important features such as factors in a factor analysis;
(c) gaining new knowledge about different populations and their mutual relations;
(d) finding characteristic set of features 'fingerprints' for the known occurrences;
(e) forming a statistical model from the above characteristic features for
(f) classifying the data set to find all similar targets.

All the cases from (a) to (f) are investigated through studying geometric properties occurring in the variable, sample and/or feature space. The techniques used for many variables is also applicable to many features and samples. Therefore, in the following sections, the term feature space (FS) is mainly applied instead of variable space.

The theoretical background to multivariate statistical methods is well treated in various textbooks (Davis 1973, Agterberg 1974, Young and Calvert 1974, Fukunaga 1990). Choosing of a suitable processing strategy is dependent on the structure of the feature space. Therefore, in this report, particular emphasis is paid to the application aspect, which is especially related to the structure visible in feature space.

In the following section some typical structures of feature spaces are shown, and suitable statistical approaches are recommended.

*Principal component analysis (PCA)*

Variables represented by maps or images may sometimes seem to be so similar that their colour combination does not show the targets of interest by different colours. The combination is therefore similar to one of the component variables and mutual correlation between the variables is high. The feature space expressed by a scatter diagram is built up of an elongated 'cloud' of observation points (Fig. 4.11). Maximal elongation, i.e. maximal variation directions of points is revealed by *principal component analysis.*

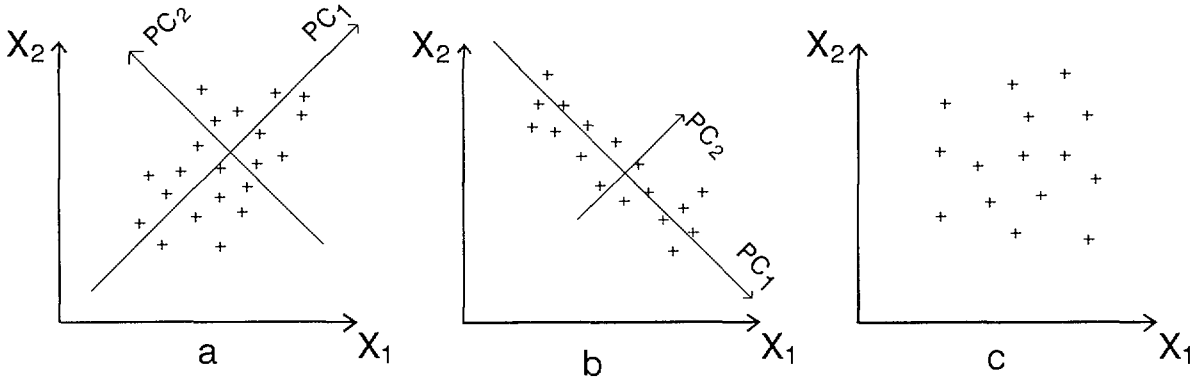



*FIG. 4.11. Principal components in scatter diagrams of variables $X_1$ and $X_2$. (a) positive high correlation, (b) negative high correlation, (c) variables not correlated and the PCs are at random.*

In the cases of high (positive or negative) correlation (Fig. 4.11 (a) and (b)) mutual contrasts of $X_1$ and $X_2$ can be enhanced by choosing new cordinate axes along and perpendicular to the major elongation of the scatter cloud. The new axes are called *principal components* PC 1 and PC 2. The principal components are orthogonal and therefore uncorrelated. The new coordinates of observations (projected onto the new axes PC 1 and PC 2) are called principal component *scores* and the correlation coefficients of the PC axes and $X_1$–$X_2$ coordinates are called principal component *loadings*. Calculation of PC:s is based on the *variance-covariance* matrix of the original variables (Davis 1973).

Principal component analysis is sometimes able to separate populations of points (see Siegal and Gillespie 1980, p. 200) expressed by clusters in FS (Fig. 4.12(a)).

Variation in the observation set may be caused by many populations, as in the case of multispectral satellite or gamma radiation data. PCA usually gives a very effective first impression of these populations (Fig. 4.13, see Plate 5). Principal component analysis can naturally be applied in more than two dimensions. This technique is in practice effectively used when there are up to 30 variables.

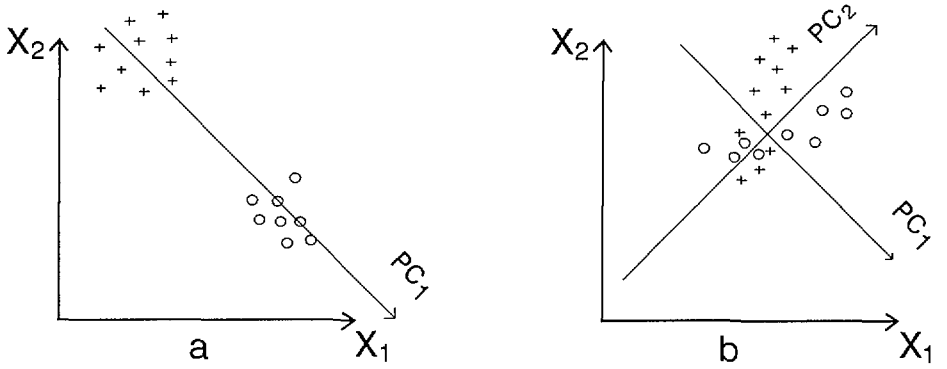



*FIG. 4.12. Principal component axes in the case of separate populations in FS. (a) PCA is applicable, (b) PCA is not able to separate clusters.*

*Factor analysis (FA)*

Just as principal component analysis was used to find uncorrelated new variables (PC:s) and therefore enhance contrasts between groups of variables, *factor analysis* is used to reduce a number of variables to a few significant uncorrelated factors which can explain the total variation in the observation set. Reducing the number of variables may often be necessary if the computer programs available are not able to handle such a large number of original variables. The original variables are standardized prior to FA and therefore the new variables, namely factors, become comparable to each other. Factor analysis of this type, when the variables are replaced by factors, is called R-mode factor analysis.

Generally, factor analysis is applicable to the same cases of feature space as for PCA (see Figs 4.11 and 4.12), mainly to study one population. FA is a powerful tool for separating derivatives within a single population, e.g. both the extreme end-members and the in-between-members. A typical example is discrimination of rock types based on chemical analyses of samples representing a magmatic differentation sequence.

The search for factors is also carried out by seeking directions of maximal variation in FS. Some statistical assumptions are however made about the nature of population from which samples were taken; namely the original random variables $X_j$ are multivariate ormally distributed and one can predict a suitable number (p) of orthogonal factors to be extracted from m original variables (p must be less than m).

The factor model (Davis 1970, pp. 500–533, Fukunaga 1990, pp. 399–417) can be expressed as:

$$X_j = \sum l_{rj} f_r^p + \epsilon_j$$

where:

$f_r$ = r'th *common factor*
$l_{rj}$ = *loading* of j'th variate on the r'th factor
$\epsilon_j$ = *unexplained variation* over variable $X_j$

If p = m and $\epsilon_j = 0$ then the problem becomes equivalent to principal component analysis.

Just as PCA utilized variance-covariance matrix to calculate new variables PC:s, factor analysis utilizes the *correlation* matrix. In factor analysis each original variable is standardized to have a mean = 0 and standard deviation = 1. Standardization transforms the former variance-covariance matrix into a correlation matrix. Eigenvectors of this matrix express elongation directions of the scatter diagram and eigenvalues show the amounts of elongation (variance) in these directions. A factor $f_r$ is a vector in the direction of an eigenvector its length being equal to square root of the eigenvalue = standard deviation in this direction. The correlation coefficients between factors and variables are called factor loadings and plots of the standardized data on the factor lines are called factor scores.

The communality of a variable is calculated as a sum of respective squared factor loadings. If communalities of each variable are near 1 this means that the factor model is appropriate, if communalities are near 0 the model is not good. It is possible to iterate the factor model by 'varimax-rotation' so that a better position for the factors and therefore better communalities are obtained.

*Q-mode* factor analysis studies the interrelations between multivariate samples instead of variables. Therefore it is an effective method for arranging a series of samples into a logical order so that their relations to each other are revealed.

Q-mode factor analysis commonly utilizes cosine theta ($\cos\theta$) matrix (instead of the correlation matrix) as the basic criteria for determination of the eigenvectors and eigenvalues and thus the factors. Q-mode factor analysis is often used in a similar way to cluster analysis, which is explained below.

Examples of using R-mode and Q-mode factor analysis in geology are given by Davis (1973) on pages 473–532.

*Clustering, discriminant analysis and classification*

*Unsupervised classification* or *cluster analysis* is used when the data structure in FS is unknown and we want to study overall grouping of the samples. Clusters can be found by studying the centers of concentration of samples, concentration gradient or correlation coefficient between samples. In the case of a small number of samples (or variables) the results of hierarchic clustering can be shown by a *dendrogram* (Fig. 4.14).

In many instances the samples cover many populations having mutually different features which occupy partly separate fields in feature spaces (Fig. 4.15).

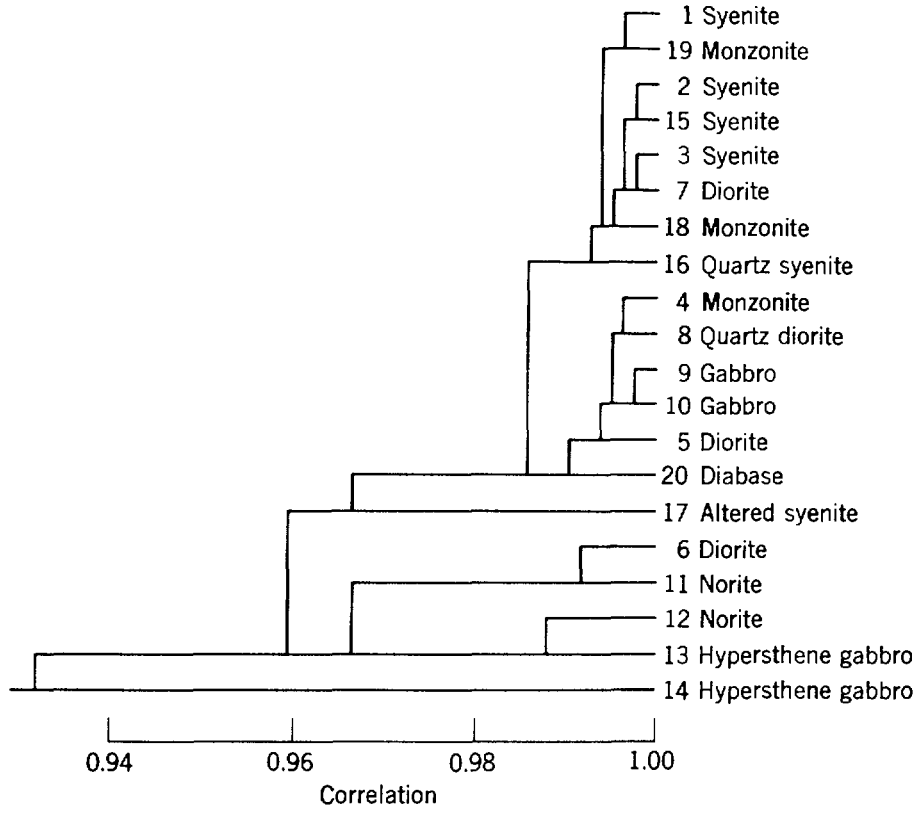



*FIG. 4.14. Results of cluster analysis for chemical analyses of igneous rocks. Dendrogram reveals the relations between different variables $X_1$–$X_{20}$. The clustering shows essentially the same arrangement as Q-mode factor analysis (from Davis 1973, Fig. 7.36).*

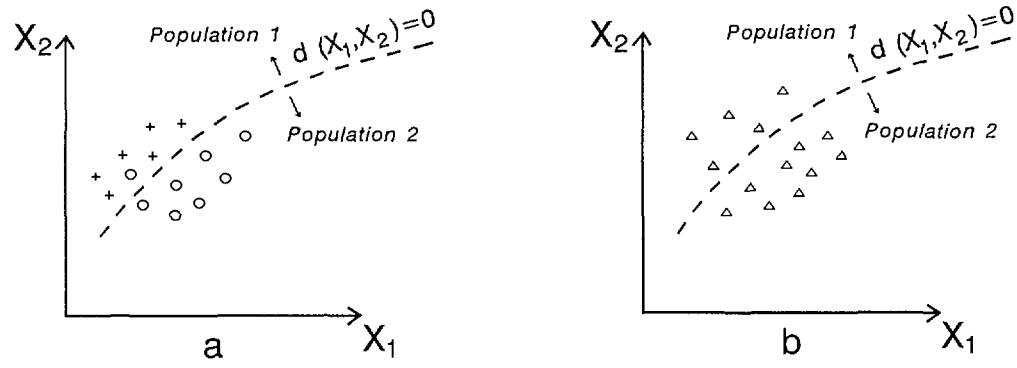



*FIG. 4.15. (a) Distribution of samples from two different model populations divided by a discriminant function $d(X_1, X_2) = 0$. (b) Additional samples from the same region are classified to respective populations 1 or 2 by the same function.*

If we know these two model distributions from past experience, we can formulate an optimal boundary $d(X_1, X_2) = 0$ between these two distributions and divide the feature space into two regions. Once the boundary is selected, we can assign a new sample without a class label to population 1 or population 2 depending on whether $d(X_1, X_2) < 0$ or $d(X_1, X_2) > 0$. We call $d(X_1, X_2)$ a *discriminant function* or a *decision boundary*. A network which detects the sign of $d(X_1, X_2)$ is called *pattern recognition network*, a *categorizer* or a *classifier*. In order to find a classifier, we must study characteristic features of each model population in our data set to find a proper discriminant function. This process is called *learning* or *training*, and the samples used to design a classifier are called *learning* or *training samples*.

What is theoretically the best classifier for given distributions? This problem is known as statistical hypothesis testing. *Bayes classifier* is the best classifier in the sense that it minimizes the classification cost. This means that Bayes classifier maximizes our money if we get 1 $ for each correctly classified sample and lose 1 $ for every misclassified sample. Bayes principle is fully explained by Fukunaga (1990).

However, implementation of an optimal Bayes classifier is often difficult because of its complexity, especially if the feature space has very many dimensions. Therefore simpler, so-called *parametric* classifiers are used, based on assumptions concerning the density functions of the desired populations or on the discriminant functions. Linear, quadratic or piecewise parametric classifiers are most commonly used (see Fukunaga 1990, pp. 124–180).

When no parametric structure can be assumed for the density functions of the populations explored, *nonparametric* techniques such as the *Parzen* and *k-nearest neighbour* techniques can be used for estimating the density functions (see Fukunaga 1990, pp. 254–297).

Traditionally however, discriminant analysis and classification are regarded as separate methods:

- The purpose of discriminant analysis is to separate all subsequent samples into previously known groups which can be separated according to $d(X_1, X_2)$ (Fig. 4.16(a)). $d(X_1, X_2)$ is fully determined by the old samples, i.e. the training samples.

- The purpose of classification — within a set of samples — is to label some samples as belonging to classes, which emerge during the classification process (Fig. 4.16(b). d ($X_1$, $X_2$) is determined through studying the structure of the FS versus distribution of the training samples, if they exist.

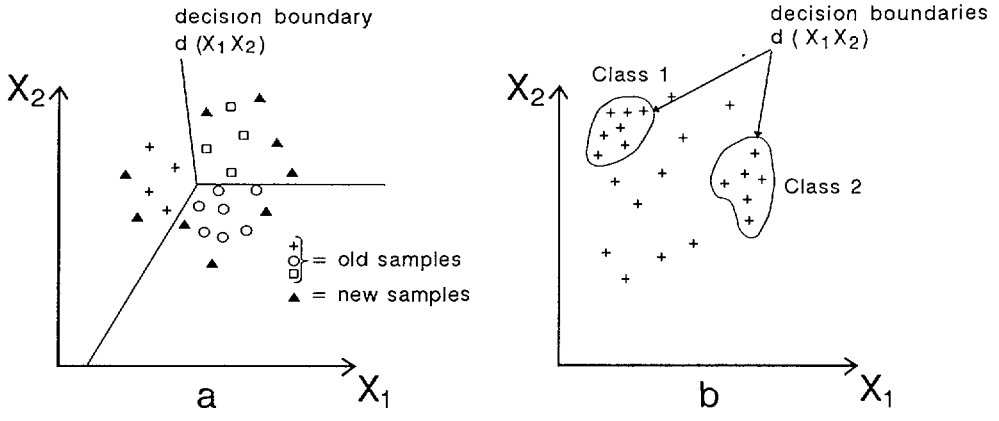



*FIG. 4.16. Difference between (a) discriminant analysis: partitioning of the whole FS into three predetermined groups and (b) classification: labelling some samples into classes emerged during classification.*

Within the scope of this report, only a few commonly used discriminant and classification methods are briefly mentioned. The interested reader may become further acquainted with the topics by reading the excellent textbooks by Krishnaiah and Kanal (1982) and Fukunaga (1990).

The discriminant function is designed to yield maximal separation of populations indicated by the training samples. Examples of *linear*, *nonlinear*, *piecewise linear* and *piecewise nonlinear* functions are illustrated in Fig. 4.17.

Classification methods are used to properly extract 'fingerprints' of desired populations (like ore deposits) from a large sample set. According to these fingerprints analogous sites for further exploration are searched for. *Supervised classification* uses training samples to identify potentially favourable classes in FS. In the geosciences the commonly used methods are *box classification*, *maximum likelihood* and *k-nearest neighbour* methods (Fig. 4.18).

In box classification the dimensions of classes (boxes) are determined by mean and standard deviation (in the direction of coordinate axes) of the training data in relation to the structure of FS. In the maximum likelihood method the classes are determined by the mean and covariance matrix of the training data in relation to the FS. Therefore the shape of a class is an ellipsoid elongated along the direction of highest variation. K-nearest neighbour classes in FS occupy space (around training data points) where distance to the nearest training point is less than a chosen constant limit. These classes are often made to adapt the FS structure according to the mutual orientation or density of the points in the FS. In these cases direction sensitive *Mahalanobis distance* and density sensitive *density gradient* are used respectively.

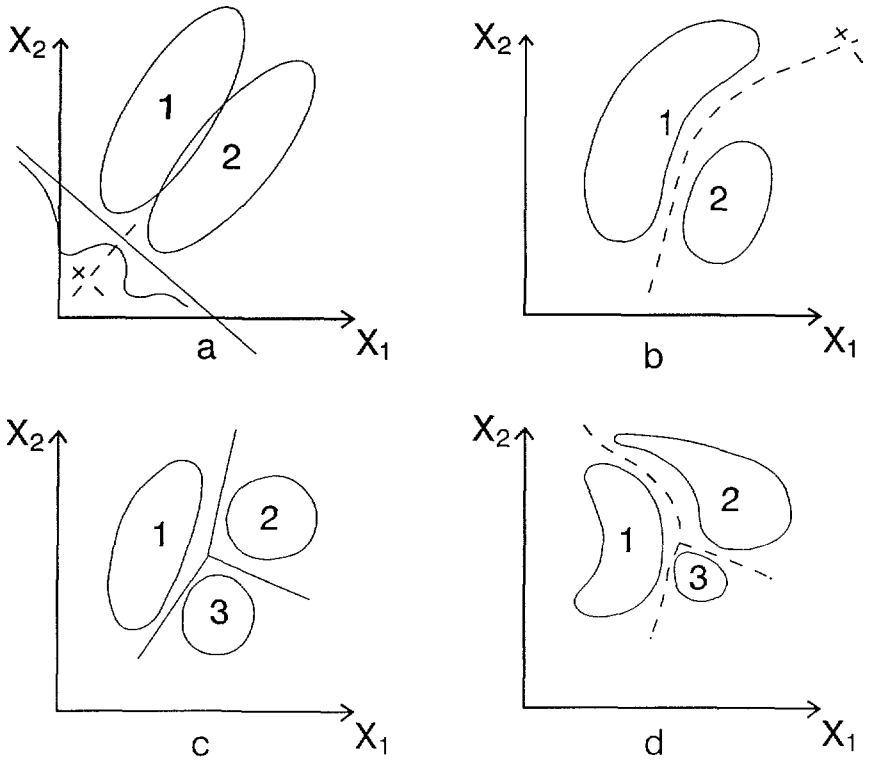



*FIG. 4.17. Different types of discriminant functions: (a) linear; (b) nonlinear; c) piecewise linear; (d) piecewise nonlinear. The distribution of observation points is described by contour lines. The numbers 1, 2 and 3 refer to different populations.*

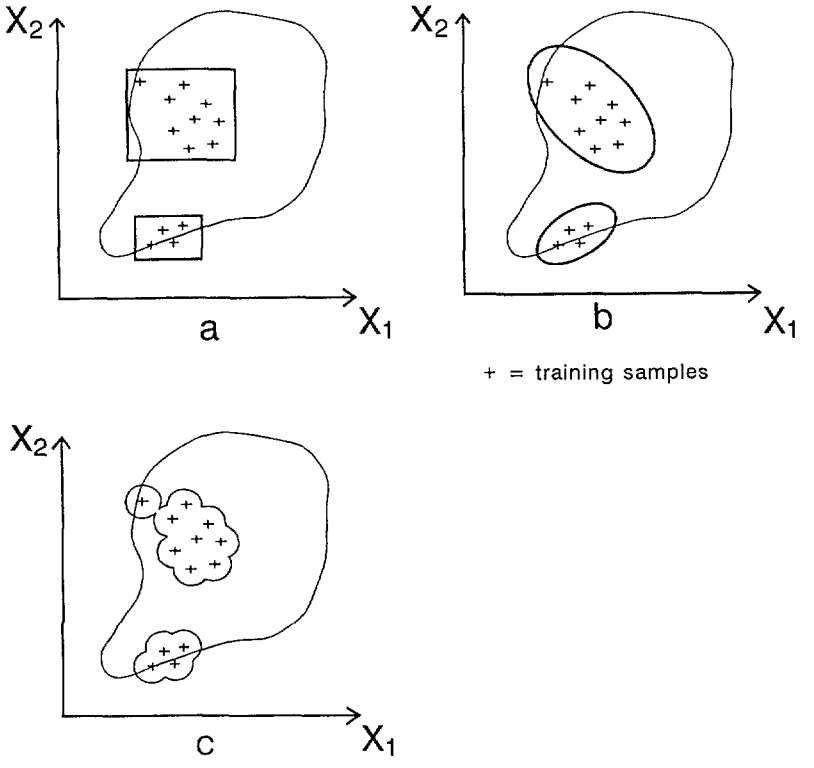



*FIG. 4.18. Three classification methods illustrated by 'favourable' classes. (a) box classification; (b) maximum likelihood ellipsoids; (c) k-nearest neighbour classes. Distribution of all samples is shown by the same density contour line in all cases.*

The final classification results may seriously differ from each other depending on the selection of the classification method. This can be easily understood when we see that the decision surfaces in each method occupy partly nonoverlapping areas in the FS (Fig. 4.18).

## 4.4. PROCESSING STRATEGIES FOR A GEOLOGIST

A processing strategy is of course dependent on the objectives of the study, nature of the input data and available finance, manpower and instrumentation. However, if our intention is to process a large data set composed of many spatial variables and training data, the processing often follows a certain strategy. In a more simple situation we may require only a part of this strategy.

The general flow chart (Fig. 4.19) can be split into a number of more detailed flow charts.

Exploration for ore deposits is always an iterative process, where attention is focused iteratively from large areas to small areas. The loop in Fig. 4.19 from "E" back to "Observations" is made in order to iterate the work from reconnaissance scale to regional and to detailed fieldwork scale.

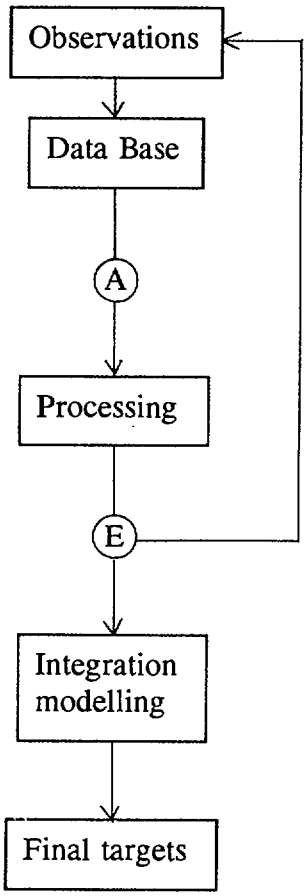



*FIG. 4.19. General flow chart for finding geological or environmental targets. A and E are the links to processing.*

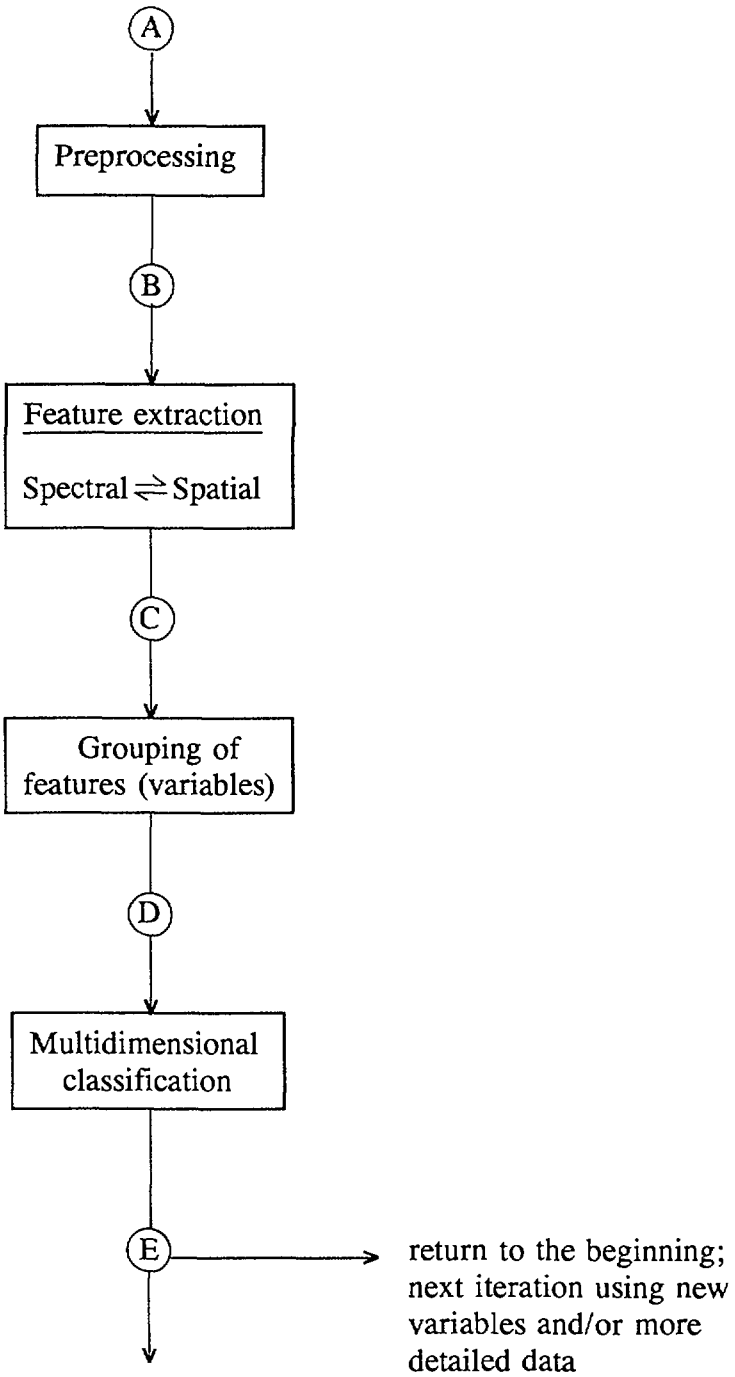



*FIG. 4.20. Main processing steps. The links A, B, C, D and E are also good points for visualization of results. The external links A and E are shown in Fig. 4.19.*

Processing involves four main steps (Fig. 4.20). The links A and E to the general flow chart are shown in Fig. 4.19.

All steps except grouping of variables have been explained in the previous chapters.

Feature extraction is a process of attempting to clarify the relations between the training data (such as ore deposits) and their characters in the spatial data (maps, images, variables) and feature spaces. It results in a reduced number of variables, significant features, and reduced — but more significant — amount of spatial variation (Fig. 4.21).

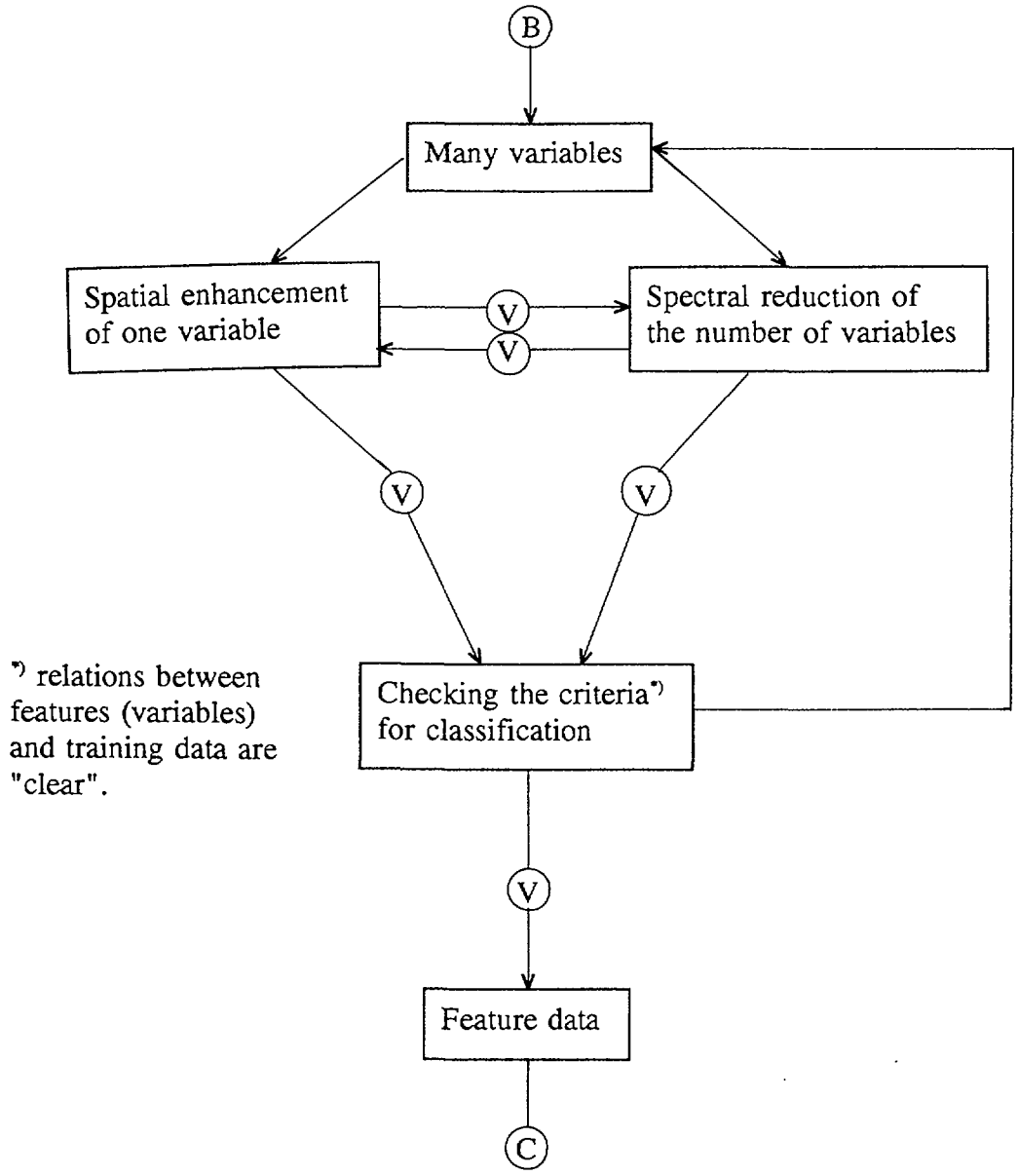



*FIG. 4.21. Spatial and spectral feature extraction A and C are external links (see Fig. 4.20) and Vs are internal links, which are all also places for visualization.*

Grouping of features and variables is extremely important. Due to geological, geophysical or hydrological, etc. factors the different indications of targets (e.g. an ore deposit) may *not* be necessarily superimposed in one geographical location but indications may occur in several places. Therefore it is necessary to enclose those features (variables) where an indication occurs in the one and same location to the same group (Fig. 4.22). Multidimensional classification is made separately for each such group.

The choice of models for multidimensional classification is often facilitated by visual inspection of the groups of features and the training data in the feature spaces (Fig. 4.23).

The results are mostly checked by *bootstrapping* (Jain et al. 1989) or a similar method in either the laboratory or in the field by checking every predicted target. Bootstrapping is a method of studying how each training datum is found by supervising the classification

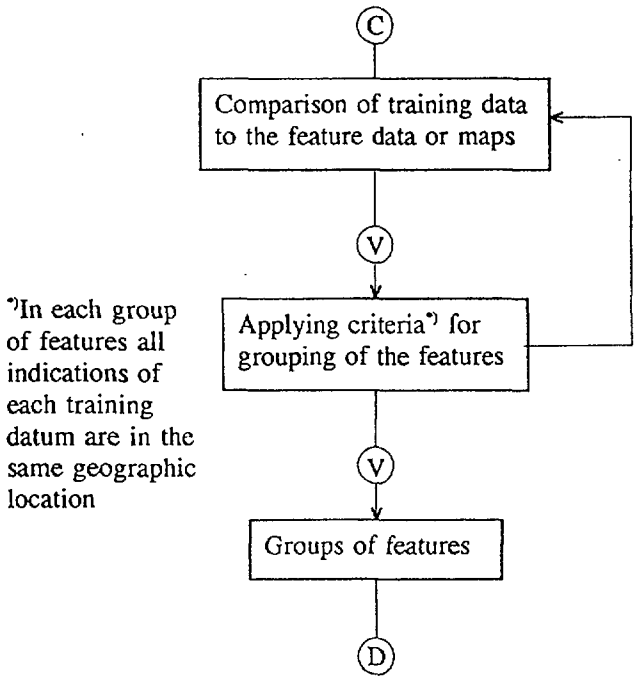



*FIG. 4.22. Grouping of features into groups of 'similar' occurrence of training data. C and D are external links (see Fig. 4.20). C, D and V also refer to visualization of the products.*

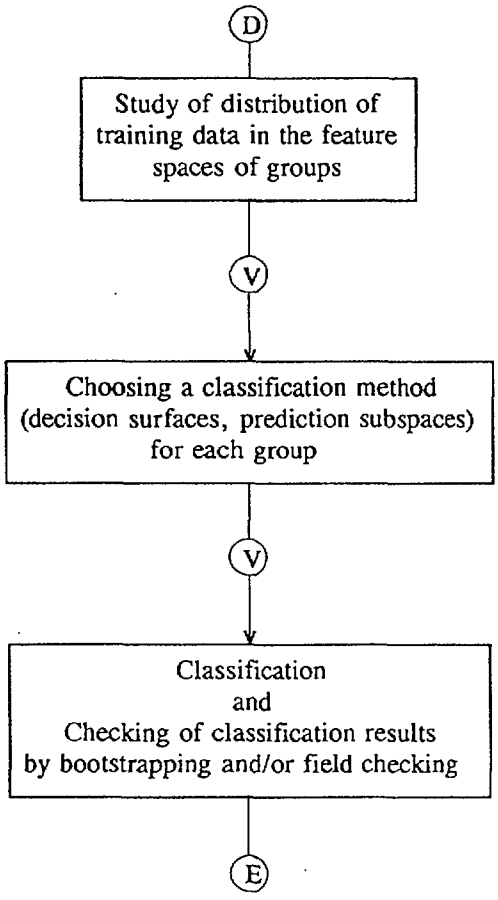



*FIG. 4.23. Different phases necessary for classification. D and E are external links (see Figs 4.19 and 4.20). D, E and V are good places for visualization.*

based on the rest of the training data (Fig. 4.23). After the critical examination of the produced classes, the results are ready for integration modelling. The classification method used does not necessarily produce 'correct' results, therefore some iteration cycles are needed. After these steps the products are ready for further integration modelling (see Fig. 4.19).

## REFERENCES


Agterberg, F.P., 1974, Geomathematics, Mathematical background and geo-science applications, Developments in geomathematics 1, Elsevier Amsterdam, 596 pp.

Banon, G.J.F. and Barrera, J., 1989, Morphological filtering for stripping correction of SPOT images. Photogrammetria, v. 43, p. 195-205.

Bernstein, R. and Fernyhough, D.G., 1975, Digital image processing: Photogrammetric Engineering, v. 41, p. 1465-1476.

Colwell, R.N., Simonett, D.S., Ulaby, F.T., Estes, J.E., Thorley, G.A. (Eds), 1983, Manual of Remote Sensing Vols. I and II, American Society of Photogrammetry.

Darnley, A. (Ed.), 1994, A Global Geochemical Database: recommendations for International Geochemical Mapping. Final Report from IGCP 259. ICSU, Paris (in preparation).

Davis, J.C., 1973, Statistics and Data Analysis in Geology, John Wiley & Sons, Inc., New York, 550 pp.

Destival, I., 1986, Mathematical morphology applied to remote sensing. Acta Astronautica, v. 13, n. 6/7, pp. 371–385.

Dougherty, E.R., Kraus, E.J. and Pelz, J.B., 1989, Image segmentation by local morphological granulometries. Proc. IGARSS'89, Vancouver, BC, Canada, July 10–14, 1989, pp. 1220–1223.

Drury, S.A., 1987, Image Interpretation in Geology, Allen & Unwin, London, 243 pp.

Drury, S.A., 1990, A Guide to Remote Sensing, Interpreting Images of the Earth, Oxford Science Publications, 199 pp.

Durand, M.A. and Flouzat, G., 1985, Quantification of the visual aspect of classified images. Photo-Interpretation, n. 3, f. 5, pp. 47–51.

Fabbri, A.G., 1984, Image Processing of Geological Data. New York, Van Nostrand Reinhold, 244 pp.

Fabbri, A.G., 1991, Spatial data analysis in raster based GIS: an introduction to geometric characterization. In, Belward, A.S. and Valenzuela, C.R. (Eds), Remote Sensing and Geographical Information Systems for Resource Management in Developing Countries. Dordrecht, Kluwer Academic Publishers, pp. 357–388.

Fabbri, A.G., van der Meer, F.D., Valenzuela, C.R. and Kushigbor, C.A., 1993, Shape analysis and multispectral classification in geological remote sensing. Mathematical Geology, in press.

Fabbri, A.G. and Kushigbor, C.A., 1989, Relationships between integrated multisensor imagery, geophysical data and geological map patterns for mineral exploration East of Bathurst Inlet, N.E.T., Canada. Proc. IGARSS'89, Vancouver, B.C., July 10–14, 1989, pp. 1414–1420.

Fahnestock, J.D., Schowengerdt, R.A., 1983, Spatially variant contrast enhancement using local range modification, Optical Engineering, Vol. 22, No. 3, pp. 378–381.

Flouzat, G. and Moueddene, K., 1986, Computer-aided interpretation of complex geological patterns in remote sensing. Proc. IGARSS'86, Zurich 8-11 Sept., 1986, ESA SP-254, pp. 783–786.

Fukunaga, K., 1990, Introduction to Statistical Pattern Recognition, Academic Press, Inc., San Diego, 591 pp.

Goetz, A.F.H., Rowan, L.C., Kingston, M.J., 1982, Mineral Identification from Orbit: Initial Results from the Shuttle Multispectral Infrared Radiometer, Science, Vol. 218, 1982, pp. 1020–1024.

Gonzales, R.C. and Wintz, P., 1987, Digital Image Processing, Addison-Wesley Publishing Company, Massachusetts, 431 pp.

Haralick, R.M., Sternberg, S.R. and Zhuang, X., 1987, Image analysis using mathematical morphology. IEEE Trans. Pattern. Anal. Machine Intell., vol. PAMI-9, n. 4, pp. 532–550.

Harris, J.L., Sr., 1977, Appl. Opt. 16(5), p. 1268.

Heiskanen, W.A., Moritz, H., 1967, Physical Geodesy, W.H. Freeman and Company, San Francisco and London, 1967, 364 pp.

Häme, T., Salli, A. and Lahti, K., 1992, Estimation of carbon storage and organic matter in Boreal forests using optical remote sensing data, in: Proceedings of the Central Symposium of the International Space Year Conference, Held in Munich, Germany, 30 March - 4 April, pp. 75–78.

IMM, The institution of Mining and Metallurgy, 1990, Papers presented at the cOnference 'Remote sensing: an operational technology for the mining and petroleum industries', IMM, London, Oct. 29–31, 1990, 285 pp.

IMSL, 1982, IMSL/IDL User's Guide, Version 1.0, 336 pp.

Jain, A.K., Dubes, R.C., Chaur-Chin Chen, 1989, Bootstrap techniques for error estimation, IEEE Transactions on Pattern Analysis and Machine Intelligence, Pami-9, 5, pp. 628–633.

Krisnaiah and Kanal, 1982.

Kruse, F.A. and Dietz, J.B., 1991, Integration of optical and microwave images for geologic mapping and resource exploration, in: Proceedings of the Eighth Thematic Conference on Geologic Remote Sensing, Exploration, Engineering, and Environment, Denver, Colorado, USA, ERIM, pp. 535–548.

Kuosmanen, V., Arkimaa, H., Gaál, G., Huhtala, T., Koistinen, E., Lindqvist, E., Murtoniemi, S., Nikander, J., Ruskeeniemi, K., Salonen, V.-P., Suppala, I., Talvitie, J., Tenhola, M., Tiainen, M. and Ward, P., 1988, Exploration target selection by integration of geodata using statistical and image processing techniques: an example from Central Finland, Part I (Text), Gaál G. (Ed.), Geological Survey of Finland, Report of Investigation 80, 156 pp.

Lillesand, T.M. and Kiefer, R.W., 1987, Remote sensing and image interpretation, John Wiley & Sons, Inc., New York, 721 pp.

Maragos, P., 1987, Tutorial on advances in morphological image processing and analysis. Optical Engineering, July 1987, v. 26, n. 7, pp. 623–632.

Maragos, P., 1989, A representation theory for morphological signal processing. IEEE Trans. on Pattern Analysis and Machine Intelligence, PAMI, v. 11, n. 6, pp. 586–599.

Maragos, P. and Schafer, R.W., 1985, A unification of linear, median, order-statistics, and morphological filters under mathematical morphology. in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, Tampa, FL, Mar. 1985, pp. 34.8.1–34.8.4, or pp. 1329–1332.

Maragos, P. and Schafer, A.W., 1987a, Morphological filters-Part I: their set theoretic analysis and relations to linear shift-invariant filters. IEEE Trans. Acoust., Speech, Signal Processing, v. ASSP-35, Aug. 1987, pp. 1153–1169.

Maragos, P. and Schafer A.W., 1987b, Morphological filters-Part II: their relations to median, order statistics, and stack filters.IEEE Trans. Acoust., Speech, Signal Processing, v. ASSP-35, Aug. 1987, pp. 1170–1184.

Martel, C., Flouzat, G., Souriau, A. and Safa, F., 1989, A morphological method of geometric analysis of images: application to the gravity anomalies of the Indian Ocean. Jour. Geophys. Res., v. 94, n. B2, pp. 1715–1726.

Matheron, G., 1975, Random Sets and Integral Geometry. New York, Wiley, 261 pp.

McDonnell, P.W., Jr., 1979, Introduction To Map Projections, Marcel Dekker, Inc., New York, 1979, p. 174.

Multala, J., 1981, The construction of gamma-ray spectrometer calibration pads, Geoexploration 19, pp. 33–46.

Paarma, H., Raevaara, H., Talvitie, J., 1968, On the interpretation of ektachrome infrared aerofilm type 8443 photographs used in mineral reconnaissance and geological surveys. Photogramm. J. Finland, Vol. 2, no. 2, pp. 1–22.

Poujade, V. and Laurore, L., 1990, Thematic segmentation method by spectral-textural cooperation using SPOT images. Photo-interpretation n. 1990-3 and 4, pp.41–46, and plates 26,27.

Roscoe, S., 1984, Assessment of mineral resource potential in the Bathurst Inlet area, NTS 76J, K, N, O including the proposed Bathurst Inlet National Park. Geol. Surv. Canada, Open File Report 788, 75 pp.

Safa, F. and Flouzat, G., 1989, Speckle removal on radar imagery with mathematical morphology. Signal Processing, April 1989, North Holland, v. 16, n. 4, pp. 319–333.

Serra, J., 1982, Image Analysis and Mathematical Morphology. New York, Academic Press, 610 pp.

# 5. INTEGRATION MODELLING

## 5.1. PREREQUISITES FOR MODELLING

Many phenomena which are observed at the earth's surface or occur at depth below the surface are of a complex nature. This is true both of processes related to man's activities and to ore forming processes. Efforts to understand these phenomena are centered usually around comparing a number of similar instances of such phenomena. They are focussed often on a few aspects of interest with the idea of seeing how they are related. Deepening understanding of these phenomena have resulted in the development of models through which this understanding can be expressed. Thus, a deepening understanding of the movement of contaminants in soils or the migration of ore-forming fluids in rocks has been achieved by the development of mathematical models based on the physics and chemistry of flow through porous media. In many instances however, the processes are of such complexity that conceptual rather than mathematical models are the most that can be realized.

In this chapter different kinds of numerical models are presented that will serve as conceptual models of certain aspects of these phenomena. For some purposes, one needs very general models which can be applied widely even though they cover few aspects of any particular situation to which they are applied. For other purposes one needs specialized models which cover more aspects of a particular situation, but which are less widely applicable. Most phenomena can be studied through several models depending upon which aspects one is interested in. In choosing an established model through which to study a particular situation, one tacitly accepts the premises on which the model was built. These will usually include some conventions concerning the way in which the model is to be applied to a particular situation.

In the examples considered in this chapter, the emphasis is to make it possible to make predictions. Two considerations need to be kept in mind when one tries to develop numerical models for this purpose. Firstly the model should be easy to work with, and secondly it should take into account all significant characteristics and fit them well. These two criteria may well act in opposition. For example, a mineral deposit model can be constructed that makes use of a small set of indicator variables whose values can be approximated in a relatively small area. However, in a survey that encompasses a large area or a region far removed from some control area, regional differences can be a significant factor and a more generalized approach may be required.

The two criteria appear not only in the choice of an established model for a particular situation, but also in the development of new models. Some aspects of a model will be devised to fit the characteristics under consideration. Others will be devised to make the model easy to work with. The implications of these two criteria for the purpose of prediction are discussed more fully in this chapter.

The most general model which we will consider involves the language of sets. Nearly all of its aspects arise from the collections of terminology used in describing phenomena on which such models are based. These models can be applied to any situation in which one is concerned about putting collections of objects together and splitting them apart in a way that characterizes the phenomena being studied and distinguishes them from other phenomena. In addition to its breadth of application these models have a second great value. They are of such generality that they can often be used as the basis for more

formal mathematical models that can be used to describe or represent specific states of nature and that form the basis for more accurate predictions. We begin our discussion with the language of sets and a perspective on the occurrence of uranium deposits in sandstones.

## 5.2. THE LANGUAGE OF SETS

One of the few statements that can be made with confidence about uranium deposits is that they compose a volumetrically insignificant part of the rocks in which they are found. Granger and Warren (1978) stated, for instance, that in roll-type deposits, the width of ore-grade uranium (>0.1 percent $U_3O_8$) in a typical roll-front deposit is commonly less than 10 m. The oxidized tongue of the roll extends 10 km or more parallel to the direction of groundwater flow. Whereas a roll-front deposit can commonly be traced for several kilometers, the tongues of oxidized rock can have an areal extent of tens of square kilometers. Such oxidized tongues make up but a small fraction of the sedimentary basins in which they are found, and these basins commonly exceed several thousand square kilometers in area. A uranium deposit makes a difficult target.

Not surprisingly, considerable attention has been given to searching for indicators that effectively increase the size of targets offered by uranium deposits. Geological factors that control the occurrence of uranium deposits ultimately will prove to be the most valuable indicators in searching for and for assessing the likelihood of occurrence of such deposits.

The basic situation is shown in Fig. 5.1. An undiscovered ore body, (A/1000), occupies an insignificant fraction of an area, A, being explored or being assessed. Surrounding the ore body is a ten times larger area, (A/100), which reflects a geochemical halo associated with the ore body. For a roll-type deposit, such an area would be represented by a zone of pyrite redeposition. Recognizing this zone would have the effect of either increasing the chances of discovery of the deposit or else increasing the credibility of a statement that an undiscovered deposit exists. Surrounding the area of mineralization is a larger area, (A/10), which represents ground favourable for the occurrence of a deposit. For roll-type deposits, this would include areas underlain by a porous, permeable, fluvial sandstone unit.

Thus, knowledge about the occurrence of roll-type uranium deposits can be used to justify broad, regional-scale, geological mapping to identify favourable areas for the occurrence of undiscovered deposits, and detailed geological and geochemical investigations within a

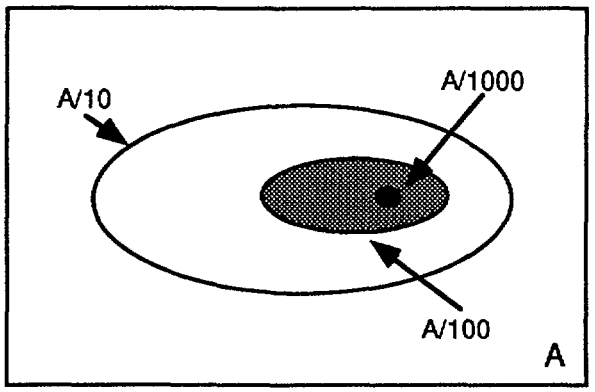



*FIG. 5.1. Diagram showing the different relative sizes of targets in exploration.*

favourable area can be used to delineate probable locations of such deposits. Similar reasoning can be applied to different types of mineral deposits. In general, diverse types of data need to be collected and organized in a way that reflects the presence of a hidden deposit. In analyzing diverse types of data, the language of sets is helpful.

The preponderence of the evidence suggests that categorical data collected in studies related to the occurrence of uranium deposits is the most valuable in terms of delineation of areas most likely to contain undiscovered deposits. Taking porosity as one example, it is necessary only to determine whether the rock can be characterized by a certain porosity or whether it lacks that porosity. In the case of rock alteration, it is essential to determine whether or not the host rock is altered and so forth. In general, recognition criteria for evaluating the favourability for undiscovered deposits rely either on the combination of a set of indicator variables whose values are above or below some threshold value or else are present or absent within an area being explored or assessed. Most often, it is the combined presence-absence of a set of variables that leads to predictors of undiscovered mineral deposits. Such combinations give rise to a logical model that makes use of the language of sets and that results in a predictive model. We turn now to a discussion of favourability as it applies to the occurrence of uranium deposits. The discussion however could just as well focus on anomalies associated with environmental hazards.

## 5.3. THE LOGIC OF FAVOURABILITY

The favourability for the occurrence of a uranium deposit can be represented as a function of a set of attributes whose combined presence (or absence) is associated with known deposits of this type. In its simplest form, each attribute can be considered as a variable whose values are represented by two mutually exclusive states, presence and absence. In exploration as in resource assessment, the task lies in the proper selection of the set of attributes used to express the favourability of occurrence. Later, we will discuss the methods for assigning weights to such attributes in order that their relative importance can be taken into account. In the present discussion, we limit attention to a single attribute (E) and its relation to the occurrence of a uranium deposit within a given area (A). For this purpose, consider Fig. 5.2 which shows a set of locations within area (A) which contains uranium deposits (D). Bear in mind that the set of locations is not meant to convey any sense of geographical location. The spatial aspects of these models will be discussed later.

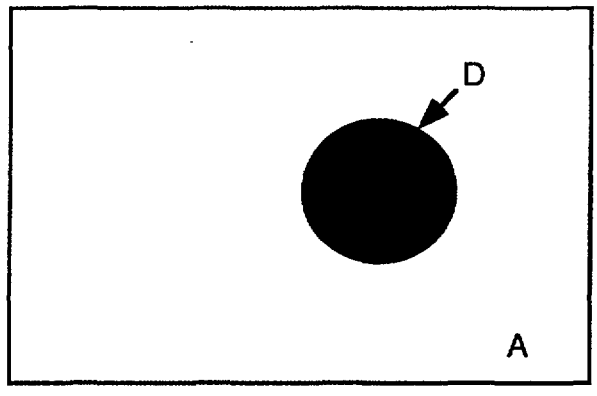



*FIG. 5.2. Diagram showing the set of locations of deposits (D) within area (A).*

The ideal situation is to choose an attribute (E) whose presence coincides with the set of locations of uranium deposits (D). This situation is shown in Fig. 5.3.

Taking A as the set of all possible locations, the relationship in Fig. 5.3 can be expressed as a statement of conditional probability

$$P\{D \mid E\} = 1,$$

that is, deposits occur if and only if E is present. In this situation, E is a necessary and sufficient condition for the occurrence of a uranium deposit. Such an attribute is a perfect discriminator and constitutes a perfect guide to undiscovered uranium deposits. For uranium deposits, the obvious candidate is natural radioactivity. Radioactive surveys have proved most valuable for locating uranium deposits at or near the surface. For uranium deposits at depth, the situation is significantly different. In these situations, a variety of types of measurements are needed. The situation is similar to that shown in Fig. 5.4 in for some cases the presence of E is associated with the occurrence of a deposit and in other cases, it is not.

For an attribute to be useful, it is desirable that $P\{D \mid E\} > P\{D \mid (\text{not } E)\}$, that is, the probability of a uranium deposit is greater when E is present than when E is absent. The extreme case occurs as shown in Fig. 5.5 in which uranium deposits occur only when E is present.

In this case, the probability is given as $P\{D \mid (\text{not } E)\} = 0$, that is, the absence of E precludes the occurrence of a deposit at the location. Such an attribute is a necessary condition for the occurrence of uranium deposits. In the case of roll-type uranium deposits, the presence of a zone of pyrite redeposition is a necessary condition for the occurrence of this type of deposit. However, not all zones of pyrite redeposition will result in the formation of a uranium deposit. The absence of such a zone precludes however the existence of such a deposit. In the search for hidden deposits in which the situation is that there are no attributes whose presence is sufficient, efforts are directed at identifying attributes whose presence is regarded as necessary for the occurrence of such deposits. In general, different attributes are required to define favourability.

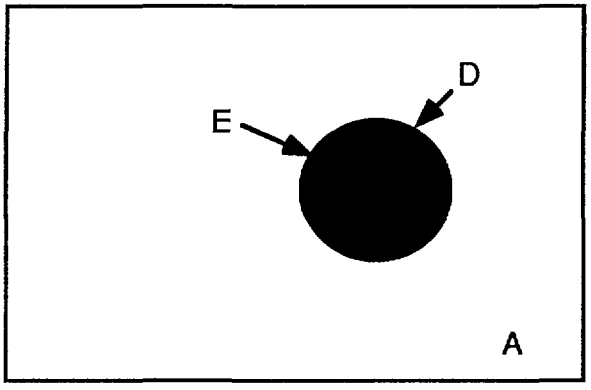



*FIG. 5.3. Diagram showing the ideal relationship between the set of uranium deposits (D) and the set of attributes (E). An occurrence of E is a necessary and sufficient condition for the occurrence of D.*

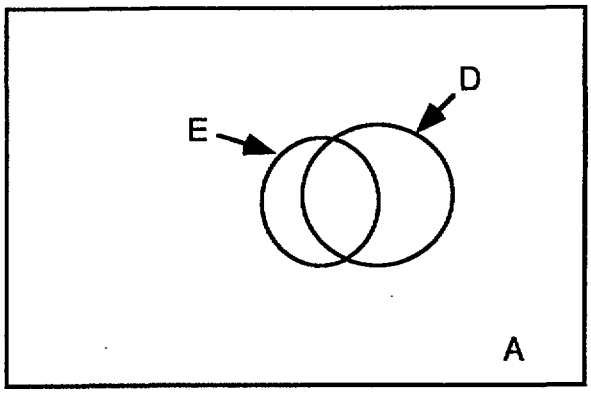



*FIG. 5.4. Diagram showing the typical relationship between the set of deposits (D) and the set of attributes (E). D occurs some of the time with E and E occurs some of the time with D.*

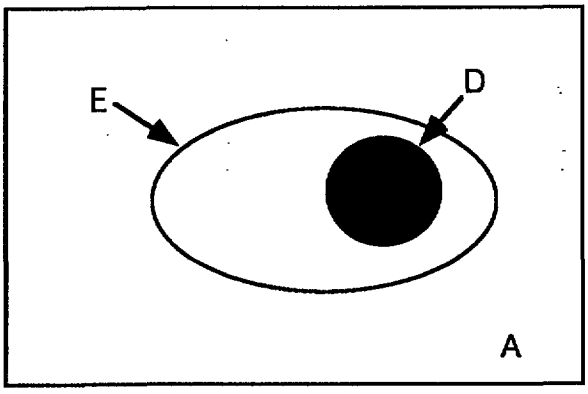



*FIG. 5.5. Diagram showing the relationship between the set of deposits (D) and the set of attributes (E). The nonoccurrence of E precludes the occurrence of D. E is a necessary condition for D.*

## 5.4. LOGICAL OPERATIONS WITH SETS

Rarely is a single attribute adequate to define the favourability of undiscovered deposits or for that matter of most phenomena. More often, it is some combination of attributes whose combined presence (or absence) is used as a measure of favourability. First, we consider logical combinations. Later, we will examine linear combinations. If now we expand the concept of an attribute (E) to represent some logical combination of variables $X_1$ and $X_2$, we can represent the state of E as defined by the logical combination shown in Table 5.1.

TABLE 5.1. DECISION RULES FOR COMBINING ATTRIBUTES WHICH DEFINES THE STATE OF ALL POSSIBLE COMBINATIONS OF PRESENCE AND ABSENCE OF VARIABLES $X_1$ AND $X_2$

| $X_1$ | $X_2$ | $X_3$ or $X_2$ | $X_1$ and $X_2$ | not $X_1$ |
|---|---|---|---|---|
| present | present | present | present | absent |
| present | absent | present | absent | absent |
| absent | present | present | absent | present |
| absent | absent | absent | absent | present |

Given these relationships, the state of any compound logical expression (E) involving more than two attributes, $X_1$, $X_2$, $X_3$, for instance, can be evaluated, as for example

$E = (X_1 \text{ or } X_2) \text{ and } (\text{not } X_3)$.

In this example, if $X_1$ and $X_2$ represent two related textural properties of a sandstone body, the presence of either of which is considered favourable for the occurrence of a uranium deposit and if $X_3$ represents a third textural property independent of $X_1$ and $X_2$, the presence of which is considered unfavourable, the above expression can be considered a host rock textural factor that is favourable only if $X_1$ or $X_2$ or both are present and $X_3$ is absent. The effect of such a logical combination is first to create a compound attribute ($X_1$ or $X_2$) that increases the degree of necessity for defining favourability and second to create a subset (not $X_3$) that increases the degree of sufficiency for defining favorablilty. The conditional probability of the combined effect is closer to 1 as compared to the conditional probability taking each attribute separately.

Using different combinations for a variety of attributes, genetic-geological models were developed for assessing the favourability for undiscovered uranium deposits in the Westwater Canyon Member of the Morrison Formation, San Juan Basin, New Mexico, USA. An example of the genetic-geological model used to assess the favourability of undiscovered trend-type uranium deposits is shown in Table 5.2.

TABLE 5.2. DECISION RULES FOR ASSIGNING FAVOURABLE STATE TO ATTRIBUTES ASSOCIATED WITH TREND TYPE URANIUM DEPOSIT MODEL

| Geological factor | Attribute | Favourable state |
|---|---|---|
| Host-rock deposition | net sandstone thickness (ft) | 200–280 |
| | sandstone/mudstone ratio | 2.5–7.0 |
| Alteration preparation | sandstone colour | grey |
| | mudstone colour | grey |
| | thickness of interval of ilmenite magnetite destruction (ft) | >0 |
| Preservation | state of oxidation | reduced |

Thus, the host-rock deposition factor was considered favourable whenever the net sandstone thickness and the sandstone/mudstone ratio fell within the stated interval; the alteration preparation factor was considered favourable whenever the sandstone colour and the mudstone colour were grey and the thickness of the interval of the ilmentite-magnetite destruction was nonzero; and the preservation factor was considered favourable when the ground was reduced. The overall favourability for uranium mineralization was represented by the combination of all three geological factors. By applying this model to a part of the San Juan basin that had been explored for trend-type deposits, an odds-ratio that compared the odds that an endowed (mineralized) or unendowed (nonmineralized) cell would be evaluated as favourable or unfavourable respectively, with the odds that an unendowed or endowed cell would be evaluated as unfavourable or favourable respectively, was calculated to be 34.1. Probability and odds are related by the expression

O = P/(1-P) where P is the probability and O is the odds. Such result is interpreted as indicating that it is 34.1 times better than if the cells were evaluated by chance. It demonstrates the predictive capability of models defined by logical combinations of attributes.

There is a large body of literature in which models defined as logical combinations of indicator variables have been applied in mineral exploration, resource assessment, and environmental studies. An excellent introduction to the principles of logic and the use of digital geographical informations systems can be found in Robinove (1986) and Ripple (1989).

## 5.5. PROBABILISTIC MODELS

As stated at the outset, a major objective using models is to be able to make predictions and the use of logical combinations of attributes in predicting the likelihood of occurrence of undiscovered deposits is one approach. What is commonly desired however is a statement that expresses the likelihood within a probabilistic framework. Such a framework is provided by the application of Bayes' theorem (Duda et al., 1978) which can be considered as an extension of the language of sets discussed earlier. Using the definition for a deposit D and attribute E, the conditional probability of a deposit occurring given the presence of the attribute can be expressed as

$$P\{D \mid E\} = P\{E \mid D\}P\{D\}/P\{E\}$$

where P{D} and P{E}, in the context described by Chung et. al. (1992), represent the probabilities of occurrence defined as the proportion of all unit areas containing a deposit D and attribute E, respectively. The left side of the above equation is referred to as the posterior probability of D given E whereas P{D} is referred to as the prior probability of D. Thus, the quantity

$$P\{E \mid D\}/P\{E\}$$

serves as a multiplier that relates the posterior probability of D to the prior probability. The greater the above quantity, the more important the attribute in predicting the occurrence of a deposit. As discussed by Chung et. al. (1992), information about P{E | D} comes from studies in and around mineral deposits whereas information about P{E} comes from regional geological mapping and studies around mineral deposits. Consider the following example contributed by Ludington and Cox (pers. comm., 1990) that involved estimating the relative importance of the age of a pluton in predicting whether the pluton has an associated W-skarn deposit. For this example, 1:500 000 scale maps of Nevada, USA. were used to estimate the total number of plutons, the number of plutons with associated W-skarn deposits, and the number of plutons that were Cretaceous in age. Of the 55 total plutons with associated W-skarn deposits, 50 were Cretaceous in age so that $P\{E \mid D\} = 50/55 = 0.91$. There were 207 total plutons so that $P\{D\} = 55/207 = 0.27$. Of the 152 plutons without associated W-skarn deposits, 67 were Cretaceous age so that $P\{E\} = (50 + 67)/207 = 0.56$. Thus,

$$P\{D \mid E\} = (0.91/0.56) \times 0.27 = 1.6 \text{ X} \times 0.27 = 0.43.$$

Knowing a pluton is Cretaceous in age raises the probability of its having an associated W-skarn deposit from 0.27 to 0.43, or 1.6 times greater than if the age of the pluton is

not considered. Similar calculations could be made for any number of other attributes that possibly would increase the probability even more. This simple example illustrates how a logic model can be elevated to a probabilitistic model. We turn now to the problem of constructing rules for prediction that are based on a large number of attributes.

## 5.6. CLASSIFICATION

In order that predictions can be made in situations that involve a large number of attributes, it is often necessary to combine the attributes into a smaller number of rules. In classification, one forms rules to decide to which class a sample belongs. The construction of prediction rules proceeds by some systematic analysis of the set of data containing values of the attributes. In constructing prediction rules there are two goals: constructing the most accurate prediction rule possible and constructing the decision rule which gives the most insight. The latter goal is important in that, in an effort to predict, the question becomes, "Which of the attributes contain significant prediction information and why?" For example, the implication is that geological factors used to define favourability are related to the processess of formation of the deposits. Such information gives insight into the results obtained when a favourability rule is applied.

A simple prediction rule that is not accurate gives misleading insights. However, in situations where the goal is to gain understanding about the phenomenon of interest, and to understand the influence of various attributes on the prediction, an easily understood and interpreted prediction rule is to be preferred to a mathematically complex rule of comparable accuracy. Thus, the two goals in constructing prediction rules are not contradictory.

### 5.6.1. The Prospector mineral consultant system

We look first at the rules of classification that were embodied in the Prospector mineral consultant system developed originally at SRI International (Duda, 1980) and subsequently modified and expanded at the US Geological Survey (McCammon, 1989). The original Prospector was an expert system designed to aid geologists in exploring for hidden mineral deposits. The system consisted of a set of rules that guided the geologist through a series of questions that led to a conclusion about the relative chances of locating a hidden deposit of a particular type at a given location. A typical session involved the geologist first describing the characteristics of a particular prospect such as the geological setting, structural controls, and kinds of rocks, minerals, and alteration products present or suspected. The system compared these observations with mineral deposit models stored in a knowledge base, noting the similarities, differences, and missing information. The system then engaged the geologist in a dialogue to obtain additional relevant information and used that information to make an assessment of the mineral potential of the prospect. The goal was to provide the geologist with advice that could normally only be obtained by consulting authorities on many different types of mineral deposits. The most noteworthy success of the system was the prediction of a hidden extension of a major copper-molybdenum deposit in eastern Washington using exploration data (Campbell et. al., 1982). To this day the system remains as one of the most successful attempts to model the task performed by geologists engaged in mineral exploration.

Within the Prospector system, the 'model' refers to a body of knowledge about a particular class of mineral deposit. The knowledge stored in the system is represented by a set of rules that are encoded in the form of an inference network. Thus, in place of a

set of attributes, there is a set of assertions (inference rules) each of which is true or false depending on the evidence that is presented to the system by the geologist. As the evidence is presented, the assertions may be definitely established, whereas others may become only more or less likely. Associated with each assertion is a probability value. The 'connections' in the inference network determine how a change in the probability of one assertion will affect those of other assertions. In general, an inference rule has the form

IF E THEN (to degree LS, LN) H,

which means "The observed evidence E suggests (to some degree) the assertion H" (hypothesis). The two parameters LS and LN establish the 'strength' of the rule and specify how the probability of H is to be updated given the existence of E. The sufficiency measure LS is defined by

$$LS = P\{E \mid H\}/P\{E \mid (\text{not } H)\}.$$

An inference rule for which LS is large means that the observation of E is encouraging for H. In the extreme case of LS approaching infinity, E is sufficient to establish H in a strict logical sense. On the other hand, if LS is much less than unity, the observation of E is discouraging for H, in as much as the observation of E diminishes the chances of H existing.

A complementary relation describes the situation in which E is known to be absent so that the necessity measure LN is defined by

$$LN = P\{(\text{not } E) \mid H\}/P\{(\text{not } E) \mid (\text{not } H)\}.$$

If LN is much less than unity, the known absence of E diminishes the chances of H existing. In the extreme case of LN approaching zero, E is logically necessary for H. On the other hand, if LN is large, the absence of E is encouraging for H. The overall structure of the Prospector model for a Western-States (USA) sandstone uranium deposit is shown in Fig. 5.6.

In Fig. 5.6, the uppermost node corresponds to the overall conclusion about the favourability of a given prospect for the occurrence of this type of deposit. The nodes below represent the major factors relevant for establishing the occurrence of this type of deposit. Each factor is established based on the evidence (observations). For example, Fig. 5.7 defines what is meant by an admissible host rock.

The evidence relevant to establishing an admissible host rock is grouped into two categories, the first pertaining to necessary conditions and the second regarding favourable conditions. The necessary conditions are defined as a logical conjunction of three factors. All three factors must be present to satisfy the necessary conditions for admissible host rock. The terminal or 'leaf' nodes in Figs 5.6 and 5.7 correspond to field evidence (observations). The evidence may consist of the presence or absence of a particular attribute or it may represent a particular value or range of values. In any case, the evidence is weighed according to its relative degree of sufficiency and necessity in establishing the next higher factor.

In general, any assertion H is either a logical combination of other statements, or is the consequent of one or more inference rules. Assuming the conditional independence of

attributes related to the next higher factor, the posterior odds O(H | E) are obtained by multiplying the prior odds by all of the likelihood ratios so that

$$O(H|E) = [\Pi_{i=1}^{n} LS_i] * O(H)$$

where O(H) = P{H}/(1-P{H}) and n represents the number of attributes related to the next higher assertion (H). If some of the $E_i$s are known to be false, $LN_i$ is substituted for $LS_i$. This can be considered as a general statement of Bayes' Rule. For example, in Fig. 5.6, the prior odds of favourable tectonic and regional conditions for the Western-States sandstone uranium deposit model is 0.0526 = 0.05/0.95. If the evidence supports the hypothesis for favourable tectonic setting, admissible host rock, and favourable host but rejects the hypothesis for favourable sedimentary tectonics, the posterior odds are calculated as 8.5 = (9* 12* 15*.1)*0.0526. Thus, the posterior probability of favourable tectonic and regional conditions becomes 0.90 = 8.5/9.5. We shall see shortly that the above equation is also the basis for the weights of evidence method of combining evidence, except that the weights of evidence is in a logarithmic form.

The design of an inference network requires the identification of the various assertions, the organization of the assertions into a hierarchical structure, the determination of the

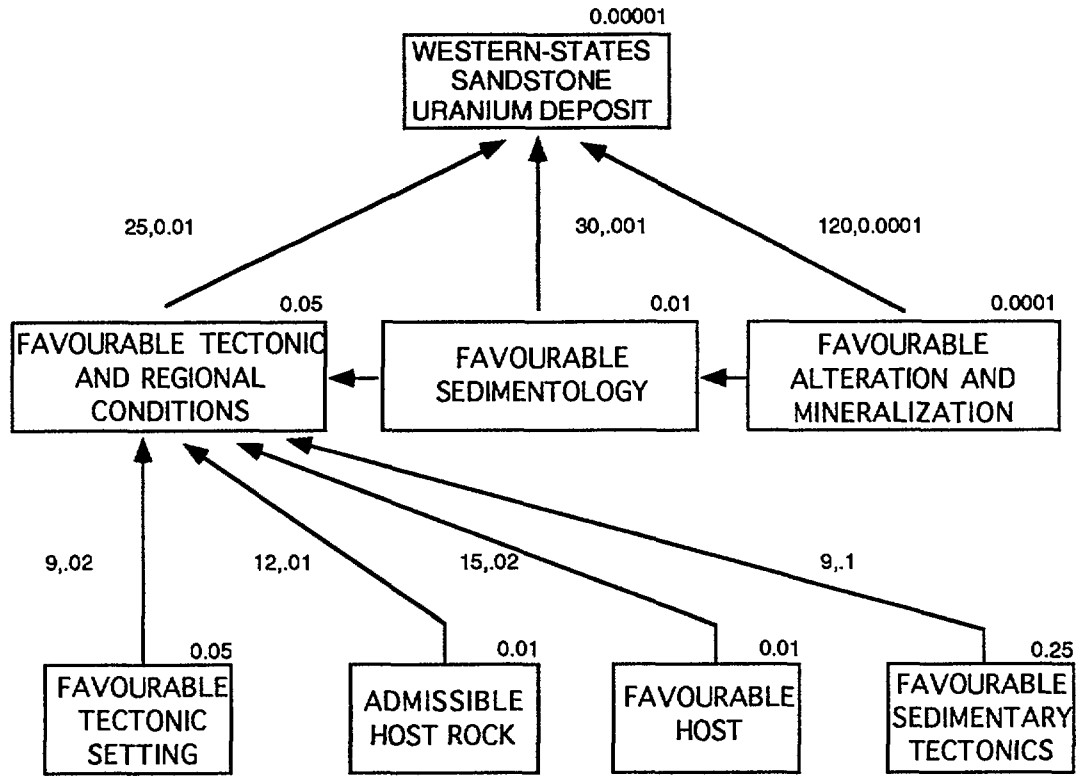
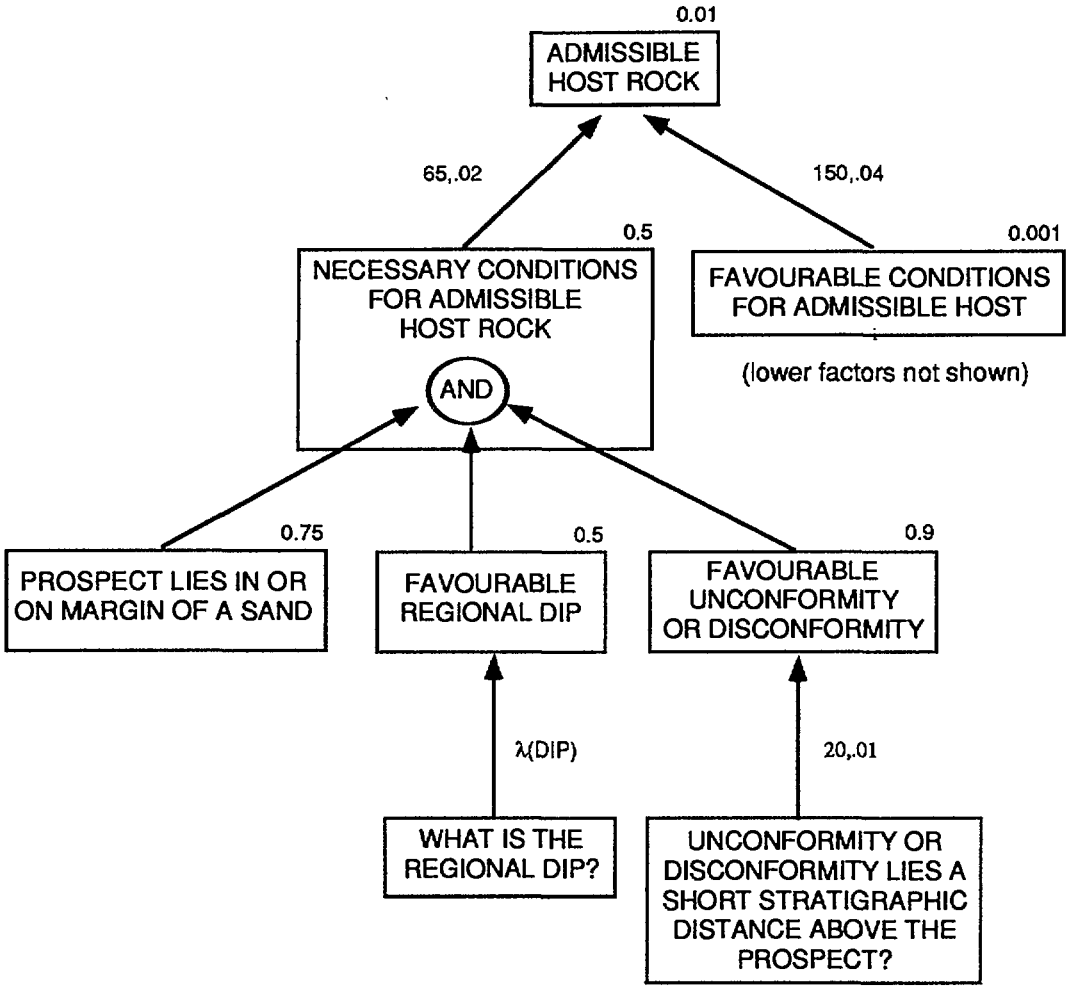



*FIG. 5.6. Inference network for the Western-States sandstone uranium deposit model after Gaschnig (1980). The prior probability and the LS, LN values are given for each factor specified in the network.*




| DIP INTERVAL (degrees) | P(INTERVAL) | λ(DIP) |
|---|---|---|
| 0-4 | 0.25 | 100 |
| 4-8 | 0.25 | 5 |
| 8-12 | 0.2 | 0.4 |
| 12-16 | 0.16 | 0.1 |
| >16 | 0.14 | 0.05 |

*FIG. 5.7. Inference network for admissible rock factor in Western-States sandstone uranium deposit model after Gaschnig (1980). The prior probability and certainty values are given for each factor specified in the network.*

types of relations that exist among assertions, and the estimation of the degrees of sufficiency and necessity of the attributes used to determine the overall favourability of occurrence of a particular deposit type. The magnitude of the task depends upon the size and complexity of the model being developed. The typical model in the Prospector system contained 60 assertions involving 80 attributes. The earlier models were designed largely for exploration. The later models were designed more for regional resource assessment. The later models required fewer assertions involving fewer attributes and were developed in a shorter time. As a final development, a model for porphyry copper was designed that was used to select drilling sites with spatial data (Duda et al., 1977; Duda et al., 1978, Katz, 1991). This development gave rise to another method applied to the problem of target selection in mineral exploration, the weights of evidence method.

### 5.6.2. The weights of evidence method

The weights of evidence method is also based on the general statement of Bayes' Rule in logarithmic form

$$\log_e O(D|E) = \sum_{i=1}^{n} w_i^k + \log_e O(D)$$

where the superscript k refers to the presence or absence of a binary pattern Ei, and

$$W_i^k = \begin{cases} W_i^+ & \text{for } _i \text{ th pattern present} \\ W_i^- & \text{for } _i \text{ th pattern absent} \\ 0 & \text{for } _i \text{ th pattern missing} \end{cases}$$

where $W_i^+ = \log_e LS$ and $W_i^- = \log_e LN$, and D is the number of deposits per unit area. The logarithmic form of the model results in the addition (rather than multiplication) of terms, and in weight factors that are positive or negative and usually less than 3. Weights defined in this manner are straight forward to interpret.

For a well-explored control area, it is assumed that a set of indicator variables are known, and are to be used as predictors of undiscovered deposits of some particular type. Furthermore, it is assumed that the locations of a number of deposits of this type are known. From these data, it is possible to calculate values of $W_i^+$and $W_i^-$ directly. The prior odds are also calculated from the data. The set of binary patterns represented by the set of indicator variables is used to generate an output pattern that can be used to predict or at least express the likelihood of occurrence of undiscovered deposits of the type being modeled. The weights of evidence are easy to interpret, simple to program, missing data can be accommodated, and patterns with complex spatial geometry can be modelled with the same computational effort as those with simple geometry.

In applying the weights of evidence method (or the Prospector method), it is assumed that the binary patterns are conditionally independent with respect to the deposits. This does not mean that the predictor pattern need be uncorrelated with each other, but that they be uncorrelated with respect to the deposits. As an extreme case, if two predictor maps are the same, as would occur if a map was replicated, then exactly the same deposits would occur where both patterns were present, and tests for conditional independence would fail. If, however, two maps were moderately well-correlated with one another (more than would be expected due to chance), deposits predicted by each pattern might well be different, and tests of conditional independence might be satisfied. In practice, some conditional dependence always is present, and judgement is required in deciding whether some patterns should be rejected from the analysis, or whether some patterns should be combined to create compound patterns (using Boolean operators) thereby circumventing the problem. An advantage to the weights of evidence method is that the weights are determined directly from the data. This can lead to the discovery of spatial relationships that were previously unknown, which may in turn lead to the recognition of new deposit models. The disadvantage is that it is not normally possible to incorporate knowledge that is not explicit in the data, although in principle there is no reason why a hybrid approach could not be applied, overriding weights determined from actual data with estimates provided by experts.

To illustrate the application of weights of evidence, consider the maps in Figs 5.8–5.13 (Plates 6–8) which show some of the layers of spatial data used in a study of gold potential in the eastern portion of the Meguma terrane, Nova Scotia, Canada. A number of small gold deposits shown in Fig. 5.8 occur within the study area. Out of 68 mineral occurrences, 35 have recorded gold production. The gold occurs in quartz veins within a sequence of early Paleozoic turbidites. The weights of evidence method was used to estimate the spatial distribution of gold potential by combining evidence from geological, geochemical and structural maps (Bonham-Carter et al, 1988; Bonham-Carter et al., 1990; Agterberg et al., 1990). The present discussion differs slightly from the earlier work in that only the 35 deposits with actual production have been used in the calculations instead of all of the occurrences and a slightly different set of maps has been used. This same set of maps will be used to illustrate the logistic regression method and fuzzy logic method to be discussed later.

The first step in weights of evidence is to select the maps to be used as evidence and to convert them to binary patterns. The second step is to combine the maps together to produce a posterior probability map that shows the spatial distribution of gold potential. The first step uses the weights of evidence to determine the binary patterns that optimize the spatial association between the points and each map pattern. The weights are then applied to the binary patterns.

Consider first the calculation of the weights for a binary pattern (ls and 0s) generated by reclassifying the 3 map units in Fig. 5.8 into 2 map units (Goldenville Formation = 1, Halifax Formation and Devonian granite = 0). Each deposit is assumed to occupy a small unit area, and all areal measurements are in terms of these units. The quantities that must be taken from the maps as input for the calculation of the weights are: (1) the area of the binary pattern (B), (2) the total number of deposits (D); (3) the number of deposits occurring on the pattern, (R); and (4) the total area (T) considered. The computing formulae for the weights are:

$$W^+ = \log_e \frac{P[B|D]}{[B|(\neg D)]} = \log_e \frac{R(T-D)}{D(B-R)}$$

$$W^- = \log_e \frac{P[(\neg B)|D]}{P(\neg B)|(\neg D)} = \log_e \frac{(D-R)(T-D)}{D(T-B-D+R)}$$

($\neg$ = not)

The OR operator takes precedence over arithmetic operators in these expressions. Taking the unit area as 1 km$^2$, the area of the Goldenville Formation (B) is 2016 units, the number of deposits within the Goldenville Formation (R) out of a total of 35 deposits (D) is 34, and the total area (T) is 2945. Inserting these numbers into the equations yields the following values for the weights: $W^+ = \log_e [34(2945-35)/35(2016-34)] = 0.355$, and $W^- = \log_e [(35-34)(2945-35)/35(2945-2016-35+34)] = -2.412$. The areas underlain by the Goldenville Formation are thus assigned a moderate weight (0.355) whereas the areas of the Halifax Formation or the Devonian granite are assigned negative weights (-2.412). The difference, $C = W^+ - W^-$, is defined as the *contrast*, equal in this case to 2.767. C is a convenient overall measure of the spatial association of the deposits with the binary pattern. The weights for $W^+$ and $W^-$ are opposite in sign except in the case when they are both equal to zero. This occurs when the number of deposits occurring within a binary

pattern is equal to the number expected due to chance, i.e., when R/D = B/T. A FORTRAN program for calculating weights together with a sample output is given in Appendix 5.A. The program includes the calculation of the standard deviations of the weights.

The prior probability used in the prediction step is usually assumed to be equal to the deposit density, or D/T. Odds (O) and probability (P) are related by the relationship O = P/(1-P). The prior log odds (also known as prior logit) is equal to $\log_e[(D/T)/(1-D/T)]$. Thus, for the Nova Scotia study area, the prior probability of a deposit in a 1 km$^2$ area is 35/2045 = 0.0119 which is equivalent to a prior logit of -4.421. Given the presence of the Goldenville Formation, the posterior logit is -4.421 + 0.355 (prior logit plus $W^+$ ) which converted to a posterior probability equals 0.0169. Given the absence of the Goldenville Formation, the posterior logit is -4.421 - 2.412 (prior logit plus $W^-$) which converted to a posterior probability equals 0.0011. Thus, the knowledge that the Goldenville Formation is present increases the probability of a deposit per 1 km$^2$ from 0.0119 to 0.0169. The knowledge that the Goldenville Formation is absent decreases the probability from 0.0119 to 0.0011. Therefore, a binary map of the Goldenville Formation of a deposit could be re-labelled a gold probability map, based on the presence or absence of a single attribute. We consider next how the weights of evidence can be used to make objective decisions in converting multi-class grey-scale maps to binary maps.

Suppose that we wish to convert the anticline map (Fig. 5.13, Plate 8) into a binary map so as to maximize the spatial association between the proximity to anticlines and the known gold deposits. Which distance should be chosen as a threshold? One approach is to calculate the weights of evidence for a series of cumulative distances, and to inspect the variation of weights and contrast as a function of distance. The results of these calculations are shown in Table 5.3.

TABLE 5.3. WEIGHTS OF EVIDENCE FOR THE 35 GOLD DEPOSITS AND THE PROXIMITY TO ANTICLINE AXES (Notice that the contrast, C (the difference between the weights), is a maximum at a distance of 1.5 km. At this distance, 29 out of 35 deposits occur, although the cumulative area of the distance buffers is only 1497 km$^2$ out of the total area of 2945 km$^2$)

area of unit cell 1 km$^2$
total area 2945 unit cells = 2945 km$^2$

| Class | area | pnts | $W^+$ | s($W^+$) | $W^-$ | s($W^-$) | C | dist (m) |
|---|---|---|---|---|---|---|---|---|
| 1 | 617 | 18 | .9150 | .2390 | -.4920 | .2430 | 1.4070 | 0-500 |
| 2 | 998 | 23 | .6726 | .2110 | -.6619 | .2896 | 1.3345 | 500-1000 |
| 3 | 1497 | 29 | .4959 | .1875 | -1.0612 | .4091 | 1.5570 | 1000-1500 |
| 4 | 1848 | 29 | .2817 | .1872 | -.7823 | .4094 | 1.0640 | 1500-2000 |
| 5 | 2133 | 30 | .1706 | .1839 | -.6633 | .4486 | .8339 | 2000-2500 |
| 6 | 2343 | 30 | .0754 | .1838 | -.3615 | .4491 | .4369 | 2500-3000 |
| 7 | 2444 | 30 | .0325 | .1837 | -.1753 | .4495 | .2078 | 3000-3500 |
| 8 | 2531 | 30 | -.0030 | .1837 | .0179 | .4499 | -.0209 | 3500-4000 |
| 9 | 2607 | 30 | -.0326 | .1836 | .2219 | .4506 | -.2546 | 4000-4500 |
| 10 | 2654 | 30 | -.0510 | .1836 | .3769 | .4511 | -.4280 | 4500-5000 |
| 11 | 2715 | 30 | -.0739 | .1836 | .6164 | .4522 | -.6903 | 5000-5500 |
| 12 | 2757 | 31 | -.0564 | .1806 | .5970 | .5054 | -.6534 | 5500-6000 |
| 13 | 2945 | 34 | | | | | | >6000 |

In Table 5.3, 29 out of 35 known deposits occur within 1.5 km of the nearest anticline, and at this distance the contrast reaches a maximum value. This suggests that a binary pattern be generated for a threshold value of 1.5 km. Notice $W^+$ is actually at a maximum for the distance class closest to the anticlines. If, however, the binary threshold is applied at this distance (500 m) the resulting binary pattern is rather small in area and only 18 out of 35 of the deposits occur in this distances. Notice that $W^-$ is at a maximum at the third distance class suggesting that beyond 1500 m, the posterior probability should be more strongly downweighted than for distances beyond 500 m. By selecting the distance at which the contrast is maximized, the combined information from $W^+$ and $W^-$ is used effectively. In practice, however, the selection of thresholds is often not so easily made as in this case, and subjective judgement must be applied to supplement the weight calculations. It should be noted that instead of using cumulative distance classes, non-cumulative distances also allow the calculation of weights, one weight per class. A problem with multi-class weights of evidence is that where only a small number of known deposits are available, the resulting weight function is unstable, that is, it fluctuates erratically for adjacent classes. In an application of multi-class weights of evidence for predicting seismicity in Western Quebec, Goodacre et al. (1993) discuss a smoothing method to overcome this problem.

By assuming that the anticlinal pattern and the Goldenville Formation pattern are conditionally independent of the known deposits, the posterior logit is calculated by adding together at each location the prior logit and the weight for each binary map ($W^+$ for presence, $W^-$ for absence).

The result is a probability map having 4 classes, corresponding to the 4 possible combinations of two binary patterns. Some of these classes will have a posterior probability greater than the prior and some will have a posterior probability less than the prior. However, if the conditional probability assumption is satisfied, the quantity (area $\times$ posterior probability) summed over the four classes can be interpreted as the predicted number of deposits and which should equal the observed number of deposits (D). The results are shown in Table 5.4.

TABLE 5.4. POSTERIOR PROBABILITIES FOR THE OVERLAP CONDITIONS OF TWO BINARY MAPS (The expected number of deposits is the posterior probability times the area. Note that the total expected number of deposits is 39, as compared to the observed number of 35, indicating that these two binary maps are not completely conditionally independent. The prior probability is 0.0119 [+ = pattern present, - = pattern not present])

| Overlap class | Geology map | Anticlines map | Area | Posterior Probability | Gold Observed | deposits Predicted |
|---|---|---|---|---|---|---|
| 1 | + | + | 1230.08 | 0.0274 | 29 | 33.70 |
| 2 | - | + | 786.73 | 0.0059 | 5 | 4.64 |
| 3 | - | - | 661.07 | 0.0004 | 1 | 0.26 |
| 4 | + | - | 267.37 | 0.0018 | 0 | 0.48 |
| Total | | | | | 35 | 39.08 |

The number of predicted deposits is 39 compared to the number of known deposits of 35. This indicates a minor problem with the assumption of conditional independence. As a consequence, the result is to slightly overestimate the posterior probabilities for locations where both patterns are present.

Using the same approach, each of the multi-class maps in Figs 5.8–5.13 (Plates 6–8) has been converted to binary form. The weights and their standard deviations are summarized in Table 5.5.

TABLE 5.5. SUMMARY TABLE OF WEIGHTS OF EVIDENCE FOR THE SIX BINARY PATTERNS

| Pattern | $W^+$ | $s(W^+)$ | $W^-$ | $s(W^-)$ | C | Name |
|---|---|---|---|---|---|---|
| 1 | 0.496 | 0.188 | -1.061 | 0.409 | 1.557 | anticlines |
| 2 | 1.127 | 0.234 | -0.589 | 0.251 | 1.717 | As, Basalm Fir |
| 3 | 0.222 | 0.210 | -0.323 | 0.290 | 0.546 | G-H contact |
| 4 | 0.355 | 0.173 | -2.412 | 1.001 | 2.767 | Geology |
| 5 | 1.133 | 0.412 | -0.857 | 0.578 | 1.989 | As, lake sed. |
| 6 | 0.376 | 0.280 | -0.170 | 0.214 | 0.546 | U/Th reatio |

| | |
|---|---|
| Prior probability | 0.01188 |
| standard deviation | 0.00200 |
| Prior log odds | -4.42065 |
| standard daviation | 0.17004 |

A posterior probability map shown in Fig. 5.14 (Plate 9) (with a maximum of $2^6$ or 64 possible overlap conditions) was generated and classified into 10 classes on the basis of area percentiles.

Regions with known deposits occur in zones with elevated posterior probabilities, as expected. For example, 8/35 deposits occur in the top 2 percent of the area, 16/35 occur in the top 5 percent of the area and 19/35 occur in the top 10 percent of the area. In addition, the deposits rated in the highest probability classes include all of the largest of the known deposits, despite the fact that the individual deposits are not weighted in the calculations. Thus, the model reflects reasonably well the distribution of the known deposits. More importantly from an exploration point of view, the map shows a number of regions where the potential to find new deposits is favourable (i.e., locations which have overlap with the indicator patterns similar to the known deposits) yet where no mineral occurrences have been reported. These target areas could be prioritized in any exploration follow-up.

There is always the possibility that some of the favourable prediction zones are 'false positives', and conversely some unrevealed deposits are not predicted. Nevertheless, the weights of evidence method serves to focus attention on possible targets in future exploration.

The weights of evidence method does not produce an estimate of the number of undiscovered deposits. Such an estimate must be made independently. If such an estimate is made, the estimate can be then used to calculate the prior probability, thereby simply increasing all the posterior probabilities. However, the rank order of areas by probability is unaffected, and it is this ranking that gives the spatial distribution of relative favourability.

### 5.6.3. The FINDER system

In the search for hidden targets such as in the case of mineral exploration, it is often useful to take into consideration the size, shape, and orientation of the targets being

sought. In this connection, a new method (FINDER) has been developed that makes use of the area of influence (Singer and Drew, 1976) and Bayesian statistics to aid in the selection of target areas on the basis of one or more variables and multiple observations (Singer and Kouda, 1988). The FINDER system employs geometric probability and the normal probability density function to integrate spatial and frequency information to produce a map of probabilities of target centers.

Missing values can be accommodated and FINDER can estimate the number of deposits. Target centers can be mineral deposits, or any other target that can be represented by a regular shape on a two dimensional map. Areas of influence include not only circular, elliptical, and annular shapes, but also sectors of annuli and rectangles with offset centres. The spatial statistics for each of the variables used to delineate the hidden targets being sought are characterized in a control area and the results applied by means of FINDER to the study area. In a test on kuroko-type deposits (Singer and Kouda, 1988), FINDER was able to identify all of the major known deposits in the Hokuroku District, Japan and suggested other favourable areas, one of which contained a discovery that was subsequently announced. More recently, FINDER was applied to the search for epithermal vein gold-silver deposits in southwest Kyushu, Japan (Singer and Kouda, 1991). A preliminary version of the FINDER program was published by Singer (1985).

### 5.6.4. The fuzzy logic method

There is a middle ground in the search for data integration models that bridge the gap between knowledge-driven and data-driven methods. Although the knowledge that is possessed by experts is not always explicit in the data, the data can be interpreted through the judgement of the expert. In large part, this can be accomplished by means of fuzzy set theory that was first systematically formulated by Zadeh (1965). A fuzzy set E is a set of ordered pairs:

$$E = \{(x, \mu_E(x)) \mid x \subset X\}$$

where X is a collection of objects and $\mu_E(x)$ is called the membership function or degree of compatibility of x in E. The range of $\mu_E(x)$ is usually defined in [0,1] where 0 expresses nonmembership and 1 a full membership. With respect to a given set of data, a membership can be assigned to each observation according to the strength of its evidence which supports some higher inference about the presence or absence of a specified target or of some physical process. Thus, a membership can be attached to an observation according to the accuracy of the data and the interpreter's expertise.

Defined this way, imprecise and incomplete information, represented using fuzzy sets, can be manipulated and processed using fuzzy set operations. One of the more robust operations for combining (aggregating) sets of attributes $E_1$, $E_2$, ..., $E_m$ due to Zimmerman and Zysno (1980) is defined as:

$$\mu_E(x) = (\Pi_{i=1}^{m} \mu_i(x))^{(1-\gamma)} (1-(\Pi_{i=1}^{m}(1-\mu_i(x)))^{\gamma} (x \subset X, 0 \leq \gamma \leq 1)$$

where

$\gamma$ is defined as an operator that is intended to balance the tendencies of overcompensating and undercompensating the effects due to any pairwise dependencies that may exist among the attributes. For instance, the use of a $\gamma$-operator applied to attributes with high and low

degrees of membership compensates for the tendency to overvalue (or undervalue) the result of combining an attribute having a high degree of membership with an attribute having a low degree of membership.

Although the concept of indicator patterns is appealing, because combining evidence in this form is relatively easy to understand and interpret, it is sometimes difficult to treat evidence in this fashion. For example, in the 'distance to anticline' map discussed for the weights of evidence method, the favourability of this factor is more likely to be some function of distance than a binary 'yes' or 'no'. One way of modelling multi-class evidence is to use fuzzy sets instead of classical sets. In a classical set, membership of the set is either yes (= 1) or no (= 0), whereas in a fuzzy set, membership can lie anywhere in the range (0, 1). Thus, if we consider as a set 'all locations favourable for gold deposits', we can construct a membership function for the anticlinal evidence which is either binary (= 1 for locations closer than 1.5 km, 0 otherwise) or fuzzy (= 1 at zero distance, and decaying to 0 at, say, 3 km). Rather than expressing fuzzy membership as a continuous mathematical function, it is usually more practical to specify discrete values as illustrated in Table 5.6. The membership functions shown in the tables reflect the subjective judgement of an exploration geologist. The functions need not necessarily increase or decrease monotonically. However, the membership values must be in the interval (0, 1). The functions are sometimes called 'possibility' functions; they should not be regarded as probabilities, because they do not satisfy the rules of probability density functions.

Having assigned membership functions to all the spatial datasets that are to be used as evidence, the next step is to combine the fuzzy evidence with fuzzy logic operators. This involves combining two or more input membership functions to produce one output membership function. Many fuzzy logic operators have been proposed, see for example the book by Zimmermann (1985). We will briefly describe five fuzzy operators, as discussed by An et al. (1991), in a paper describing the use of fuzzy logic for predicting base metal and iron deposits by combining geological geophysical datasets.

The fuzzy OR is simply the maximum of the input membership functions. Thus at a particular location, the class of each of the evidence maps is used to 'look-up' the corresponding membership function values from tables, and whichever map has the maximum function value determines the value of the output membership function. For example, given a location underlain by the Halifax Formation and where the lake sediment As is class 4, then the membership function of the two-map combination is max {0.25,0.50} = 0.50. Any number of maps can be combined simultaneously. The disadvantage of this operator is that any single location in which only one dominant piece of evidence prevails, there is no 'increased' effect of two or more favourable pieces of evidence occurring together.

The fuzzy AND is similar to OR, except that the minimum is used instead of the maximum. Thus at any one location the output is controlled by the 18 January 1994 presence of the least favourable evidence, which is usually unsatisfactory. For the same example above, the output is min {0.25, 0.50} = 0.25.

The fuzzy algebraic product is simply the product of the membership function values. For the example above, the product is (0.25 × 0.50) = 0.125. The fuzzy algebraic sum is defined as

sum = 1 - product (1-membership function).

TABLE 5.6. FUZZY MEMBERSHIP FUNCTIONS FOR NOVA SCOTIA GOLD DEPOSITS

Proximity to anticlinal axes map

| Map | Class | Legend | Membership |
|---|---|---|---|
| Anticlines | 1 | <0.5 km | 0.75 |
| | 2 | 0.5-1.0 | 0.70 |
| | 3 | 1.0-1.5 | 0.65 |
| | 4 | 1.5-2.0 | 0.60 |
| | 5 | 2.0-2.5 | 0.55 |
| | 6 | 2.5-3.0 | 0.50 |
| | 7 | 3.0-3.5 | 0.40 |
| | 8 | 3.5-4.0 | 0.30 |
| | 9 | 4.0-4.5 | 0.20 |
| | 10 | 4.5-5.0 | 0.10 |
| | 11 | >5.0km | 0.05 |

Geological map

| Map | Class | Legend | Membership |
|---|---|---|---|
| Geology | 1 | Goldenville Formation | 0.95 |
| | 2 | Halifax Formation | 0.25 |
| | 3 | Devonian Granite | 0.05 |

Airborne radiometrical map

| Map | Class | Legend | Membership |
|---|---|---|---|
| U/Th ratio | 1 | High | 0.05 |
| | 2 | . | 0.05 |
| | 3 | . | 0.05 |
| | 4 | . | 0.10 |
| | 5 | . | 0.30 |
| | 6 | . | 0.50 |
| | 7 | Low | 0.60 |

Biogeochemical map

| Map | Class | Legend | Membership |
|---|---|---|---|
| As-Balsam Fir | 1 | >73.5ppm | 0.85 |
| | 2 | 16.0-73.5 | 0.80 |
| | 3 | 7.3-16.0 | 0.75 |
| | 4 | 5.2-7.3 | 0.70 |
| | 5 | 3.9-5.2 | 0.60 |
| | 6 | 3.2-3.9 | 0.40 |
| | 7 | 2.8-3.2 | 0.20 |
| | 9 | 2.4-2.8 | 0.10 |
| | 10 | 0.3-2.4 | 0.05 |

TABLE 5.6. (cont.)

Lake sediment geochemical map

| Map | Class | Legend | Membership |
|---|---|---|---|
| As-Lake seds. | 1 | >73.5-166 ppm | 0.80 |
| | 2 | 112-142 | 0.75 |
| | 3 | 52-112 | 0.70 |
| | 4 | 28-52 | 0.50 |
| | 5 | 17-28 | 0.30 |
| | 6 | 12-17 | 0.20 |
| | 7 | 7-12 | 0.10 |
| | 8 | 2-7 | 0.05 |

Proximity to contact between Goldenville and Halifax formations

| Map | Class | Legend | Membership |
|---|---|---|---|
| Distance to contact | 1 | <0.5 km | 0.50 |
| | 2 | 0.5-1.0 | 0.45 |
| | 3 | 1.0-1.5 | 0.40 |
| | 4 | 1.5-2.0 | 0.35 |
| | 5 | 2.0-2.5 | 0.20 |
| | 6 | 2.5-3.0 | 0.10 |
| | 7 | 3.0-3.5 | 0.05 |
| | 8 | 3.5-4.0 | 0.01 |
| | 9 | 4.0-4.5 | 0.01 |
| | 10 | 4.5-5.0 | 0.01 |
| | 11 | >5.0 | 0.01 |

For the two map example, sum = 1-(1-0.25)(1-0.50) = 0.625. Neither the fuzzy product nor the fuzzy sum are very satisfactory on their own, but they can be combined with a fuzzy gamma operator to produce a compromise result.

The gamma operation is defined as the algebraic product raised to the power $(1-\gamma)$ multiplied by the algebraic sum raised to the power $\gamma$, where $\gamma$ is a parameter in the range (0, 1). For the example above, and selecting $\gamma = 0.95$, the output is $(0.125^{0.05})(.625^{0.95}) = 0.5767$. An et al. (1991) discuss the choice of suitable gamma values. The results of applying the gamma operation are appealing, because they produce a reasonable compromise between conflicting sources of evidence.

Figure 5.15 (Plate 9) shows the result of applying the gamma operation to the Nova Scotia datasets, followed by classifying the result into 10 classes as shown. Surprisingly, even by choosing arbitrary membership functions, the results look reasonable, with the major known deposits being satisfactorily predicted.

One of the advantages of this approach is that several alterative sets of membership functions, possibly generated by a variety of exploration geologists with conflicting views, can be tried out by simply adding new columns to the membership function tables. The resulting predictive maps can then be compared and the sensitivity of the results to varieties of opinion assessed.

### 5.6.5. The Dempster-Shafer method

An interesting and powerful knowledge-driven approach for combining map evidence uses the idea of belief functions and the Dempster-Shafer rule of combination. The idea is that instead of estimating a single belief function (superficially similar in concept to fuzzy membership functions), two functions are estimated for each evidence map, one being a lower bound of belief and the other an upper bound. The interval between the functions is called an 'evidential interval' and represents the uncertainty or lack of precise knowledge about the evidence. This means that P {E} does not neccessarily equal 1-P{(not E)}, except when the evidential interval shrinks to zero, in which case the Dempster-Shafer method is the same as the Bayesian method. The approach has been used by Moon (1990), (Moon et al., 1991) and (Chung and Moon, 1991) for mineral exploration. The advantage of this approach is that it explicitly handles the uncertainty in estimating the membership function. Like the fuzzy membership functions, belief functions are usually estimated subjectively by a mineral deposit expert. The output using this method is not a single predictive map, but a pair of maps, one showing the lower bound, the other the upper bound, thereby directly incorporating uncertainty into the results.

## 5.7. REGRESSION

In regression, one forms rules to predict some numerical value. The construction of regression prediction proceeds by some systematic analysis of the set of data containing values of the attributes together with the related values of the attributes to be used in the prediction rule. Regression differs from classification in that one wants to predict some numerical value, and not a class. Such a prediction for instance might be the probability of occurrence or it might be the undiscovered endowment or it might be the value of some environmental variable. Whatever the purpose, the standard approach is through the use of linear regression in some form. In this section, we will consider a number of different approaches to the problem of prediction using regression methods. We consider first a method for estimating undiscovered metal endowment within large areas. The results are based on a presumed relationship between a numerical measure of geological favourability and the spatial distribution of metal endowment.

### 5.7.1. One-level prediction

The method was first developed for estimating the undiscovered uranium endowment in the San Juan basin, New Mexico (McCammon et al., 1986) as a part of the National Uranium Resource Evaluation (NURE) program conducted by the Department of Energy from 1973 to 1980 (US Department of Energy, 1980). The method came about largely because the extensive amount of data collected as part of this program dictated that the consolidation of such a large quantity of information into a succinct evaluation of the potential uranium resources be explained in an explicit, traceable manner. For the San Juan basin study, it was possible to grid all of the data that were collected, namely, the lithologic and petrographic characteristics of the main host rock unit, the locations of the exploratory drilling, and the discovered uranium endowment (defined as the quantity of $U_3O_8$ contained in material with a grade of at least 0.01 percent $U_3O_8$). Using these data, it was possible to relate the geological favourability with the discovered uranium endowment in the explored portion of the basin, and using this relationship, to estimate the undiscovered uranium endowment in the remaining unexplored portion of the basin.

In applying the method, it is assumed the quantity of geological information is sufficient to allow calculation of (1) a numerical measure of favourability, (2) the extent of exploration, and (3) the discovered endowment for a suitably defined grid of equal area cells (for the San Juan basin study, the area was divided into square cells approximately 4 km on a side). In this context a favourability measure is defined to be some calculatable function of the data. High values are associated with the occurrence of deposits of the type represented by the model under consideration and low values are associated with nonoccurrence of such deposits.

Metal endowment within the unexplored portion of a favourable cell (i.e., a cell for which the favourability measure is greater than a specified favourability threshold level) is estimated to be proportional to the discovered endowment found within a suitably chosen control area which has been explored. It is assumed that the geological factors that control the occurrence and distribution of deposits throughout the area being assessed (the study area) are the same as the factors that control the occurrence and distribution of deposits within the control area Because each unexplored cell is determined to be either favourable (endowed) or unfavourable (unendowed), the procedure has been named one-level prediction.

A fixed endowment to be assigned to a unit area of the unexplored, favourable portion of the model area is calculated by dividing the total discovered endowment within the control area by the area of the explored, favourable portion of the control area. Each unexplored or partially explored cell in the model area is classified as favourable if the favourability measure is greater than or equal to the selected favourability threshold and is classified as unfavourable otherwise. The estimate of undiscovered endowment in a favourable cell is the product of the calculated fixed endowment times the area of the cell (or the area of the unexplored portion of a partially explored cell). Cells determined to be unfavourable are estimated to contain no endowment. The estimate of the total undiscovered endowment in the model area is then the sum of those endowments assigned to unexplored or partially explored cells.

The measure of the error generated when the fixed endowment is assigned to a unit of unexplored, favourable area is defined as the absolute value of the difference between errors due to overestimating and to underestimating the endowment per unit area in the control area. If the favourability threshold level is set too high, the method tends to underestimate the true endowment whereas if the threshold level is set too low, the method tends to overestimate true endowment. Using a method best described as trial and error, a threshold level is selected that tends to minimize these two errors. The result is a per unit or area-normalized error measure. To estimate the error for the calculation of the total undiscovered endowment in the model area, the area-normalized error measure must be adjusted to reflect the size of the area being assessed.

One-level prediction was used to estimate the undiscovered uranium endowment in the San Juan basin, New Mexico (McCammon et al.,1986). Discovered uranium endowment data were obtained from uranium properties, published reports, news releases, company stockholder reports, and unpublished information. In order that estimates of the discovered endowment could be made for each cell, orebody plan view maps were superposed on a grid and the tonnage portion of each orebody was assigned to the cell according to the area covered by the orebody. Information on the extent of exploration was derived from orebody and drillhole location maps compiled from published and unpublished sources. A value of the fraction of area explored was assigned to each cell

based on the number and distribution of drill holes and the area underlain by orebodies. Stratigraphic data for the main host rock unit, the Westwater Canyon Member of the Morrison Formation, were obtained from 1800 geophysical logs and 40 measured outcrop sections. Petrographic data were obtained by microscopic examination of chip samples from 40 petroleum test wells, core descriptions of 9 test holes, 50 company exploration logs, and 14 measured surface sections.

These data and knowledge gained from past studies of uranium deposits in the San Juan basin were used to select five separate control areas on the basis of the different deposit types identified and on the extent of exploration in and around areas of known uranium deposits. The geological characteristics associated with each cell were assigned a level of favourability with respect to each of the five deposit models used in the study. For each of the models, a favourability threshold level was selected and estimates of undiscovered uranium and estimation error were calculated. These estimates were summed to yield the estimate of the total undiscovered uranium endowment and the estimated total error.

Using the one-level prediction method, the total undiscovered uranium endowment for the main host rock unit, the Westwater Canyon Member, was estimated at 2.6 million tonnes $U_3O_8$ with an estimated error of 0.25 million tonnes $U_3O_8$. This estimate was roughly twice that obtained by the U.S. Department of Energy (1980) in the National Uranium Resource Evaluation (NURE) program but similar to results from the assessment by Harris and Carrigan (1981), who used a method different from the method of the Department of Energy. The strengths and weaknesses for all three methods are discussed in Harris (1984).

An artificial example is provided which contains a variety of combinations of inputs for a given set of cells. The data are shown in Table 5.7.

In this example, the area to be assessed has been divided into 25 cells. Of these 25 cells, 19 are defined as being within the model area and 6 are defined as being outside the model area. Cells within the model area are designated by m, the model area indicator function (i.e., m = 1).

Of the 19 model area cells, 5 comprise the control area. Cells within the control area are designated by v, the control area indicator function (i.e., v = 1). Some of the cells within the control area have discovered endowment (i.e., cells for which d, the discovered endowment, is greater than zero) whereas other cells (i.e., d = 0) do not. However, all cells in the control area have been explored to some extent (i.e. p, the fraction explored, is greater than zero). In general, cells defined as the control area will be explored to a considerable extent and usually, but not always, will contain discovered endowment.

In applying the one-level prediction method, it is assumed that all cells within the model area can be assigned a degree of favourability. However favourability is defined, each cell within the model area shown in Table 5.7 has a declared value of f, which represents the degree of favourability. As stated earlier, higher values of f are more likely to be associated with the occurrence of deposits of the type represented by the model whereas lower values are more likely to be associated with the nonoccurrence (absence) of deposits of the type represented by the model.

A judgement needs to be made about the threshold level for f. The proper choice and the methods used to make this judgement are beyond the scope of this example. For our

TABLE 5.7. ARTIFICIAL EXAMPLE DATA FOR ONE-LEVEL PREDICTION
(Variable symbols for the table are: m = model area indicator function, v = control area indicator function, p = fraction explored, f = favourability, and d = discovered endowment)

| | | Column | | | | |
|---|---|---|---|---|---|---|
| Row | Variable | 1 | 2 | 3 | 4 | 5 |
| 5 | m | 0 | 1 | 1 | 1 | 1 |
| | v | 0 | 0 | 0 | 0 | 0 |
| | p | 0 | 0 | 0 | 0 | 0 |
| | f | 0 | 2 | 4 | 5 | 6 |
| | d | 0 | 0 | 0 | 0 | 0 |
| 4 | m | 1 | 1 | 1 | 1 | 1 |
| | v | 0 | 0 | 0 | 0 | 0 |
| | p | 0 | 0 | 0.2 | 0.2 | 1 |
| | f | 2 | 3 | 4 | 2 | 5 |
| | d | 0 | 0 | 200 | 100 | 400 |
| 3 | m | 1 | 1 | 1 | 1 | 0 |
| | v | 0 | 1 | 1 | 1 | 0 |
| | p | 0 | 0.2 | 0.8 | 0.8 | 0 |
| | f | 3 | 4 | 2 | 5 | 0 |
| | d | 0 | 0 | 200 | 300 | 0 |
| 2 | m | 1 | 1 | 1 | 1 | 0 |
| | v | 0 | 0 | 1 | 1 | 0 |
| | p | 0 | 0.2 | 0.2 | 1 | 0 |
| | f | 4 | 2 | 3 | 5 | 0 |
| | d | 0 | 0 | 0 | 100 | 0 |
| 1 | m | 0 | 1 | 1 | 0 | 0 |
| | v | 0 | 0 | 0 | 0 | 0 |
| | p | 0 | 0 | 0.2 | 0 | 0 |
| | f | 0 | 1 | 3 | 0 | 0 |
| | d | 0 | 0 | 0 | 0 | 0 |

purpose, the judgement is made that a favourability greater than or equal to 3 defines the threshold value.

To assist in the calculation of the results obtained in the one-level prediction method, a QuickBasic subroutine is provided in the Appendix 5.B. Comments within the subroutine explain the calculations to be performed in order to determine the calibration constant, the total endowment, and the error associated with the estimate of the endowment.

The calculated results for the artificial example in Table 5.7 are shown in Table 5.8.

For a threshold value of 3, the calculated calibration constant is 272.7. This value represents the average endowment per favourable cell within the control area. The calculated area normalized error measure is 36.4.

For the model area, the calculated total undiscovered endowment is 2563.6, the total endowment is 3863.6, and the calculated error associated with the calculated undiscovered endowment is 237.3.

TABLE 5.8. CALCULATED ONE-LEVEL PREDICTION RESULTS FOR DATA IN TABLE 5.7

| One-level prediction | Value |
|---|---|
| Favourability threshold | 3 |
| Total size of control area | 5 |
| Size of explored part of control area | 3 |
| Size of favourable part of control area | 2.2 |
| Discovered endowment in control area | 600 |
| Calibration constant | 272.7 |
| Total size of model area | 19 |
| Size of explored portion | 4.8 |
| Size of unexplored portion | 14.2 |
| Total discovered endowment | 1300 |
| Total undiscovered endowment | 2563.6 |
| Total endowment | 3863.6 |
| Area-normalized error measure | 36.44 |
| Calculated error | 237.3 |

In the above example, it was assumed that all cells within the model area could be assigned a degree of favourability and that each cell within the model area value, f, represented the degree of favourability. In general, the favourability is defined as some type of function that relates the observations (evidence) to a particular model. As pointed out earlier, the evidence is related to the factors that define such a model. Typically, favourability is defined as a weighted linear combination of factors, $f_i$, each of which contributes information about the presence (or absence) of some stage of the model. For a model involving n factors, the favourability is expressed as

$$f = a_1f_1 + a_2f_2 + \ldots + a_nf_n$$

For a given set of $f_i$'s defined by a particular model which can be evaluated for a set of cells within a control area, it is possible to determine the weights, $a_i$, in such a way that the $f_i$'s match the degree of favourability (assigned) of each of the cells in the control area. The weights, $a_i$, are determined by solving the matrix equation

$$(X'X)\ a = \lambda,a$$

where X' is the transpose of X and where $\lambda$ is the largest eigenvalue of the product matrix (X'X). X is the m $\times$ n matrix of variables for m control cells. For cells outside the control area for which observations (evidence) are made, the favourability can be calculated using the linear model defined above. As an example of this type of model, Table 5.9 shows the characteristic weights applied to the attributes that were shown in Table 5.2.

In this instance, the weights were calculated using the data for 5 control cells such that for each cell, each attribute was assigned a value +1 if favourable, and -1 otherwise. The weights refer to attributes that have been transformed in a similar manner. Because the weights are nearly equal, the favourability of a cell outside the control area can be determined simply by adding the number of attributes that are judged to be favourable. This reduces to a model in which favourability is defined by the number of elements in a set.

TABLE 5.9. CALCULATED CHARACTERISTIC WEIGHTS FOR ATTRIBUTES ASSOCIATED WITH TREND TYPE URANIUM DEPOSIT MODEL

| Geological factor | Attribute | Characteristic weight |
|---|---|---|
| Host-rock deposition | net sandstone thickness (ft) | 0.17 |
| | sandstone/mudstone ratio | 0.17 |
| Alteration preparation | sandstone colour | 0.17 |
| | mudstone colour | 0.17 |
| | thickness of interval of ilmenitemagnetite destruction | 0.15 |
| Preservation | state of oxidation | 0.17 |

### 5.7.2. Logistic regression

In some sense, favourability is a measure of probability. Thus, for a group of cells for which some cells contain mineral deposits and some cells do not, and there are data available for all cells, a model which has received widespread attention is the logistic regression model defined by

$$f = e^{(a_1f_1+a_2f_2+\ldots+a_nf_n)}/(1+e^{(a_1f_1+a_2f_2+\ldots+a_nf_n)})$$

where f represents a probability between 0 and 1 and the $a_i$'s represent the weights applied to the data represented by the $f_i$'s. The estimates of the weights can be obtained using nonlinear methods. A logistic model therefore makes it possible to determine the probability of occurrence directly given a set of data in a control area. Such a model can then be applied to cells outside the control area.

As the name suggests, the logistic regression model also deals with log odds or logits. Both logistic regression and weights of evidence are examples of log-linear models. However, the logistic model involves the calculation of only one coefficient for each predictor variable, plus a constant, instead of a pair of weights. Instead of modelling the relationship between each predictor pattern and the deposit pattern independendy, logistic regression involves the simultaneous determination of all the coefficients. No assumption of conditional independence is required. In practice, the logistic coefficients may be difficult to interpret on an individual basis, a problem similar to regression analysis in general.

Table 5.10 shows the logistic coefficients for the six binary predictor maps used in the weights of evidence methods for the Nova Scotia example.

In Fig. 5.16 (Plate 10), the probability of a gold deposit per unit area is shown as a map, based on the logistic equation. The probability values have been classified into the same area percentile classes as used in Fig. 5.14 (Plate 9), allowing direct comparison of the two methods. As can be seen from Fig. 5.17 (Plate 10), the differences between the two predictions are minor.

TABLE 5.10. SUMMARY OF LOGISTIC COEFFICIENTS FOR THE SIX BINARY MAP PATTERNS

| Pattern | Coefficient | St. dev. | Name |
|---|---|---|---|
| 0 | -8.22 | 1.08 | constant |
| 1 | 1.28 | 0.46 | anticlines |
| 2 | 1.54 | 0.35 | As, Balsam Fir |
| 3 | 0.62 | 0.37 | Contact between Goldenville and Halifax Formation |
| 4 | 2.19 | 1.03 | Geology |
| 5 | 0.11 | 0.47 | As, lake sediments |
| 6 | 0.24 | 0.36 | U/Th ratio |

Both logistic regression and weights of evidence are useful methods. Weights of evidence is easier to interpret, and is invaluable for deciding how to generate the binary patterns. Logistic regression is valuable as a check on the consequences of conditional dependence. Reddy et al. (1991) discuss the application of the logistic model to predicting base metal potential in a greenstone belt in Manitoba, Canada.

## 5.8. LOCAL SPATIAL MODELLING

This section deals primarily with the study of locally spatially referenced data and associated statistical models and processes. It may seem that it is sufficient to build a proper physical model, in which the data are independent observations from Poisson random variables with means determined by a particular transform of a true underlying map. Unfortunately, in many situations, the inverse problem of inferring the true intensities is illposed for this to provide a satisfactory solution. This is particularly true in environmental studies that involve such intensities associated with nonpoint source pollution or yearly point source outputs. In these types of studies, a stochastic model for the true underlying structure is needed that is at once globally flexible, yet locally constrained to produce severe discontinuities only when there is convincing evidence of their existence in the data.

Historically, the analysis methods of local spatial modelling have been limited by sparse sampling. Modern data acquisition methods now have greatly circumvented this sampling limitation. The overall result has been the development of databases that provide a high spatial resolution and a synoptic realization of a given process under study. The challenge in the use of these newer data sources is to summarize eloquently and to increase understanding of enormous quantities of information. There are many open research problems in the area of visual data-analytic techniques. For example, how does one display the uncertainty connected with contour lines on a statistical map, and what is an effective way of displaying more than one spatially expressed variable? Part of the answer to this question lies in the geostatistical analysis of spatial data.

Like most of the other methods that have been discussed in this chapter, geostatistics is mostly concerned with local spatial prediction. Much of the emphasis is based on the spatial prediction method known as kriging, a term coined by Matheron (1963) in honour of D.G. Krige, a South African mining engineer.

What sets geostatistics apart from the methods discussed earlier in the chapter is the notion of dependence among the data that is defined by the distance between successive sampling points. Specifically, suppose that the differences of variables lagged some distance (h) vary in a way that depends only on h so that

$$\text{var}\,(Z(s+h) - Z(h)) = 2\gamma(h) \text{ for all } s,\ s + h$$

where the spatial index s is 2- or 3-dimensional. The function $2\gamma(h)$, which is a function only of the difference between the spatial locations s and s+h, has been called the variogram by Matheron (1963).

The computation, interpretation, and modelling of variograms form the basis of many studies of small areas in ore reserve estimation and pollution prediction. In effect, the variogram model is used to define the spatial correlation structure and controls the way that kriging weights are assigned to sample data that are used to interpret the particular phenemenon of interest. A large literature exists on fitting variogram models based on various assumptions about the underlying spatial structure. It is a subject which is beyond the scope of this chapter and we refer the reader to the literature for the details. Two recent references that provide an excellent introduction to geostatistics are Isaaks and Srivatstava (1989) and Cressie (1991).

## 5.9. SUMMARY

In this chapter, we examined a variety of models for integrating data collected in mineral exploration, resource assessment, and environmental studies. The primary aim in using these methods is to make credible predictions based on the data. The path to more accurate and reliable prediction in our opinion may lie in the blending of expert knowledge and the data. There is a large body of knowledge that exists outside the data and it is this knowledge that provides the basic framework upon which credible predictions can be made. Expert knowledge provides us with the rules; the data provide us with the evidence. The optimum strategy in our opinion is to devise methods for transforming the data in a way that reflects the true nature of the underlying phenomena. Such transformations must be consistent with the models that are used to represent the phenomena being investigated.

Some methods for transforming data include (a) probability measures, (b) favourability measures, and (c) membership functions. For each method, a separate set of assumptions is required. Each set of assumptions depends upon the particular goals of the investigation. Whatever method is used, a basic requirement is that the dependencies among the data are accounted for. Such a requirement is necessary in order to avoid bias in the results. Bias can be reduced or eliminated by the use of logical combinations, statistical correlation, or by reformulation of the model. In all cases, every effort should be made to remove potential bias.

In this chapter, we have avoided discussing the actual implementation of methods into a GIS. Most GISs provide some capability for modelling, although this is often limited. Many GISs provide a modelling language, specifically designed for combining maps (in raster or vector mode) with arithmetic, Boolean and other operators. The language may also permit the use of conditional statements, reference to attribute tables, and the use of user-defined functions. Of the methods described in this chapter, it will certainly be possible to program Boolean and fuzzy logic operations. It may also be possible to set up

Prospector-type inference networks and weights of evidence models. It may even be possible to program models using Dempster-Shafer, one-level prediction and logistic regression. A popular alterative approach for such specialized models is to use the GIS for data compilation and for the construction of maps, then to export ASCII files in a convenient data structure to an external software package, and finally return the results to the GIS for further display and analysis. This approach offers a measure of flexibility in that use is made of the GIS for ease of data input, transformation and display, yet permits specialized modelling software to be developed independently of a particular GIS's requirements.

## REFERENCES


Agterberg, F.P., 1989, Computer programs for mineral exploration: Science v. 245, pp. 76–81.

Agterberg, F.P., Bonham-Carter, G.F., and Wright, D.F., 1990, Statistical pattern integration for mineral exploration, in Gaal, G. and Merriam, D.F., Eds, Computer Applications in Resource Estimation Prediction and Assessment for Metals and Petroleum, Pergamon Press, Oxford-New York, pp. 1–21.

An, P., Moon, W.M., and Rencz, A., 1991, Application of fuzzy set theory for integration of geological, geophysical and remote sensing data: Canadian Journal of Exploration Geophysics, v.27, pp. 1–11.

Bonham-Carter, G.F., Agterberg, F.P., and Wright, D.F., 1988, Integration of geological datasets for gold exploration in Nova Scotia: Photogrammetry and Remote Sensing, v. 54, n. 11, pp. 1585–1592.

Bonham-Carter, G.F., Agterberg, F.P., and Wright, D.F., 1990, Weights of evidence modelling: A new approach to mapping mineral potential, in F.P. Agterberg and G.F.Bonham-Carter, Eds, Statistical applications in the earth sciences: Geological Survey of Canada Paper 89, pp. 171–183.

Campbell, A.N. Hollister, V.F., Duda, R.O., and Hart, P.E., 1982, Recognition of a hidden mineral deposit by an artificial intelligence program: Science v. 217, pp. 927–929.

Chung, C.F. and Agterberg, F.P., 1980, Regression models for estimating mineral resources from geological map data: Mathematical Geology, v. 12, pp. 473–47B.

Chung, C.J. and Moon, W. M., 1991, Combination rules of spatial geoscience data for mineral exploration: Geoinformatics, v. 2, n. 2, pp. 159–169.

Chung, C.F., Jefferson, C.W., and Singer, D.A., 1992, A quantitative link among mineral deposit modeling, geoscience mapping, and exploration-resource assessment: Economic Geology, v. 87, p. 194–197.

Cressie, N., 1991, Geostatistical analysis of spatial data, in Spatial Statistics and Digital Image Analysis: National Academy Press, Washington, D.C., pp. 87–108.

Duda, R.O., 1980, The Prospector system for mineral exploration: Final Report SRI Project 8172, SRI International, Menlo Park, California.

Duda, R.O., Hart, P.E., Nilsson, N.J., Reboh, R., Slocum, J., and Sutherland, G.L., 1977, Development of a computer-based consultant for mineral exploration: Annual Report SRI Projects 5821 and 6415, SRI International, Menlo Park, California, 202 pp.

Duda, R.O., Hart, P.E., Barrett, P., Gaschnig, J.F., Konolige, K., Reboh, R., and Slocum, J., 1978, Development of a computer-based consultant for mineral exploration: Annual Report SRI Projects 5821 and 6415, SRI International, Menlo Park, California, 193 pp.

Duda, R.O., Gaschnig, J., and Hart, P., 1979, Model design in the Prospector consultant system for mineral exploration, in Donald Michie, ed., Expert systems in the microelectronic age: University of Edinburgh Press, Edinburgh, pp. 153–167.

Gaschnig, J., 1980, Development of uranium exploration models for the Prospector consultant system: Final Report SRI Project 7856, SRI International, Menlo Park, California, 603 pp.

Goodacre, A., Bonham-Carter, G.F., and Agterberg, F.P., 1993, Statistical association of seismicity with drainage patterns and magnetic anomalies in Western Quebec: Tectonophysics, v.217, p. 285–305.

Granger, H.C. and Warren, 1978, Some speculations on the general geochemistry and hydrology of roll-type uranium deposits, in Resources of the Wind River Basin Wyoming Geological Association Annual Field Conference, Casper, Wyoming, pp. 349–361.

Isaaks, E.H. and Srivastava, R.M., 1989, An introduction to applied statistics: Oxford University Press, New York, 561 pp.

Harris, D.P., 1984, Mineral resources appraisal:Oxford University Press, New York, 445 pp.

Harris, D.P. and Carrigan, F.J., 1981, Estimation of uranium endowment by subjective geological analysis – A comparison of methods and estimates for the San Juan basin, New Mexico: Economic Geology, v. 76, n. 5, pp. 1032–1055.

Hirst, K.E. and Rhodes, F., 1971. Conceptual models in mathematics, in David Wheeler, Ed., Mathematical Studies no. 5.: George Allen and Unwin Ltd., London, 182 pp.

Katz, S.S., 1991, Emulating the Prospector expert system with a raster GIS: Computers & Geosciences, v. 17, n. 7, pp. 1033–1050.

Konolige, K., 1979, Bayesian methods for updating probabilities, in R.O. Duda, P.E. Hart, Knolige, and R. Reboh, A computer-based consultant for mineral exploration, Final Report SRI Project 6415, SRI International, Menlo Park, pp. 83–146.

Matheron, G., 1963, Principles of geostatistics: Economic Geology, v. 58, pp. 1246–1266.

McCammon, R.B., 1980, Research on uranium resource models, A progress report: Part II, Geologic decision analysis and its application to genetic-geologic models: US Geological Survey Open-File Report 81–556, 18 pp.

McCammon, R.B., 1989, Prospector II: The Annual AI Systems in Government Conference Proceedings, March 27-31, 1989, Washington, D.C., pp. 88–92.

McCammon, R.B., Finch, W.I., Kork, J.O., and Bridges, N.J., 1986, Estimation of uranium endowment in the Westwater Canyon Member, Morrison Formation, San Juan Basin, New Mexico, using a data-directed numerical method, in C.E. Turner-Peterson and E.S. Santos, eds., A basin analysis case study – the Morrison Formation, Grants uranium region: American Association of Petroleum Geologists, Studies in Geology No. 22, pp. 331–355.

McCammon, R.B. and Kork, J.O., 1992, One level prediction--A numerical method for estimating undiscovered metal endowment: Nonrenewable Resources, v. 1, n. 2, pp. 139–147.

Moon, W.M., 1990, Integration of geophysical and geological data using evidential belief function: IEEE Transactions, Geoscience and Remote Sensing, v. 28, pp. 711–720.

Moon, W. M. and An, P., 1991, Integration of geophysical, geological and remote sensing data using fuzzy set theory: Geoinformatics, v. 2, n. 2, pp. 171–176.

Moon, W.M., Chung, C F. and An, P., 1991, Representation and integration of geological, geophysical and remote sensing data, Geoinformatics, v. 2, n. 2, pp. 177–188.

Reddy. R.K., Agterberg, F.P., and Bonham-Carter, G.F., 1991, Application of GIS-based logistic models to base-metal potential mapping in Snow Lake area, Manitoba: Proceedings Canadian Conference on GIS, Ottawa, Canada, March 18-22, 1991, pp. 607-618.

Ripple, W.J., Ed., Fundamentals of geographic information systems: A compendium: American Society for Photogrammetry and Remote Sensing, Falls Church, Virginia, 248 pp.

Robinove, C.J., 1986, Principles of logic and the use of digital geographical information systems: U.S. Survey Circular 977, 19 pp.

Singer, D.A., 1985, Preliminary version of FINDER, a Pascal program for locating mineral deposits with spatial information: US Geological Survey Open-File Report 85–590, 24 pp.

Singer, D.A.and Drew, L.J., 1976, The area of influence of an exploratory hole: Economic Geology, v. 71, pp. 642–647.

Singer, D.A. and Kouda, R., 1988, Integrating spatial and frequency information in the search for Kuroko deposits of the Kokuroku District, Japan: Economic Geology, v. 83, pp. 18–29.

Singer, D.A. and Kouda, R., 1991, Application of the FINDER system to the search for epithermal vein gold-silver deposits: Kushikino, Japan, a case study: Geoinformatics, v. 2, n. 2, pp. 113–123.

US Department of Energy, 198(), An assessment report on uranium in the United States of America: U.S. Department of Energy GJO-111(80), 150 pp.

Zadeh, L.A., 1965, Fuzzy sets: Information and Control, v. 8, pp. 338–353.

Zimmermann, H.J., 1985, Fuzzy Set Theory and Its Applications: Kluwer-Nijhoff Publishing, Boston, Dordrecht, Lancaster, 363 pp.

Zimmermann, H J. and Zysno, P., 1980, Latent connectives, human decision making: fuzzy sets and systems, v. 4, pp. 37–51.

## Appendix 5.A
## FORTRAN PROGRAM FOR CALCULATING WEIGHTS OF EVIDENCE

```
      program odds
      character*1 q
 1    print *, ' area of study region ?'
      read *, s
      print *, ' area of binary map pattern ?'
      read *, b
      print *, ' area of unit cell?'
      read *, unit
      print *, ' no of deposits on pattern?'
      read *, db
      print *, ' total no of deposits ?'
      read *, ds
      if (db.le.ds) go to 6
       print *, 'error '
      stop
 6    s=s/unit
      b=b/unit
      pbd=db/ds
      pbdb=(b-db)/(s-ds)
      wp=alog(pbd/pbdb)
       vp=1./db+1./(b-db)
      sp=sqrt(vp)
      pbbd=(ds-db)/ds
      pbbdb=(s-b-ds+db)/(s-ds)
      wm=alog(pbbd/pbbdb)
       vm=1./(ds-db)+1./(s-b-ds+db)
      sm=sqrt(vm)
      c=wp-wm
      sc=sqrt(vp+vm)
      priorp=ds/s
      vprip=priorp/s
      sprip=sqrt(vprip)
      sprilo=sprip/priorp
      prioro=priorp/(1-priorp)
      prilo=alog(prioro)
      cpp=exp(prilo+wp)
      cpp=cpp/(1.+cpp)
      cpm=exp(prilo+wm)
      cpm=cpm/(1.+cpm)
        print 2, db,ds,b,s,wp,sp,wm,sm,c,sc,c/sc,priorp,sprip,prilo,
     +sprilo,cpp,cpm
 2    format(//////' Number of deposits on pattern       ',f10.2/
     +          ' Total number of deposits            ',f10.2/
     +          ' Area of binary pattern (unit cells) ',f10.2/
     +          ' Total area (unit cells)             ',f10.2/
     +          ' W+                                ',f15.4/
     +          ' Standard deviation of W+          ',f10.4/
     +          ' W-                              ',f15.4/
     +          ' Standard deviation of W-          ',f10.4/
     +          ' Contrast                        ',f15.4/
     +          ' Standard deviation of Contrast    ',f10.4/
     +          ' Contrast/Standard deviation       ',f10.4/
     +          ' Prior probability               ',f15.4/
     +          ' Standard deviation of prior prob  ',f10.4/
     +          ' Prior log odds                  ',f10.4/
```

```
     +              ' Standard deviation of prior log odds ',f10.4/
     +              ' Cond prob of deposit given pattern   ',f15.4/
     +              ' Cond prob of deposit given no pattern',f15.4///)
      print *, 'another case ?'
      read(*,'(a)') q
       if (q.eq.'Y'.or.q.eq.'y') go to 1
     stop
     end
```

**Example output**

```
Number of deposits on pattern                34.00
Total number of deposits                     35.00
Area of binary pattern (unit cells)        2016.00
Total area (unit cells)                    2945.00
W+                                           .3551
Standard deviation of W+                     .1730
W-                                         -2.4125
Standard deviation of W-                     1.0005
Contrast                                     2.7675
Standard deviation of Contrast               1.0154
Contrast/Standard deviation                  2.7256
Prior probability                            .0119
Standard deviation of prior prob             .0020
Prior log odds                             -4.4206
Standard deviation of prior log odds         .1690
Cond prob of deposit given pattern           .0169
Cond prob of deposit given no pattern        .0011
```

**Appendix 5.B**
**A QUICKBASIC SUBROUTINE FOR CALCULATING**
**THE RESULTS OBTAINED IN THE ONE-LEVEL PREDICTION METHOD**

```
SUB onelvl (r%, c%, m%(), v%(), p(), d(), f(), fc, c, di, ud, x, fv, xv, xu, vt, a, e)
'-------------------------------------------------------------------------------------------------
'       Input data set specifications
'       Data for the calculations consist of four parameters specified for
'       each rectangular cell in the model, which is assumed to be
'       a subset of a rectangular array of cells.
'           r%, c%   - number of rows and columns in array of cells
'           m%(i,j)  - model cell indicator
'               (0 = not in model, <>0 = model cell)
'           v%(i,j) - an indicator variable specifying the control
'               area (0 = not control area, <>0 = control area)
'           p(i,j)   - the fraction of the cell that has been explored
'               [0 <= p(i,j) <= 1]
'           d(i,j)   - the discovered endowment in the cell
'               [d(i,j) >= 0]
'           f(i,j)   - the cell favorability index for the cell
'           fc       - favorability threshold


'-------------------------------------------------------------------------------------------------

'Description of output
'           c    - calibration constant
'           di   - discovered endowment
'           ud - undiscovered endowment
'           x    - size of the explored portion
'           fv   - size of favorable portion in control area
'           xv - size of the explored portion of the control area
'           xu   - size of the unexplored portion
'           vt   - size of control area
'           a    - area normalized error measure
'           e    - prediction error measure for the model area


'-------------------------------------------------------------------------------------------------
'zero the sums
        di = 0: ud = 0: x = 0: a = 0:  xv = 0: xu = 0: vt =0: vd = 0: fv = 0
'loop through the rectangular array and accumulate sums
        FOR i% = 1 TO r%
            FOR j% = 1 TO c%
                IF m%(i%, j%) THEN
                    ifv% = (f(i%, j%) >= fc)
                    vt = vt + v%(i%, j%)
                    di = di + d(i%, j%)
                    ud = ud - ifv% * (1 - p(i%, j%))
                    vd = vd + v%(i%, j%) * d(i%, j%)
                    xv = xv + v%(i%, j%) * p(i%, j%)
                    x = x + p(i%, j%)
                    fv = fv - ifv% * v%(i%, j%) * p(i%, j%)
                    xu = xu + 1! - p(i%, j%)
                    a = a + ifv% * v%(i%, j%) * p(i%, j%) * (d(i%,j%) <= 0)
                    a = a - (NOT (ifv%)) * v%(i%, j%) * p(i%, j%) * (d(i%, j%) > 0)
                END IF
```

```
            NEXT j%
        NEXT i%
'calculate the calibration constant, undiscovered resources,
'area-normalized error measure, and prediction error measure
        IF fv > 0 THEN
            c = vd / fv
            ud = c * ud
            a = c * ABS(a) / xv
            e = SQR(xu * xv) * ABS(a)
        END IF

END SUB
```

# 6. TECHNICAL REQUIREMENTS OF GIS PROCESSING

## 6.1. ROLE OF GIS IN EARTH AND ENVIRONMENTAL SCIENCES

It is a fact that the natural sciences are devoted to the study of phenomena which are only partly described by exact observations and which are associated to various levels of uncertainty. For this reason, we should not be surprised that the GIS, both as a cartographic tool and as an instrument of decision support, is having a seemingly destabilizing effect for institutions with a mandate of producing maps. The very traditional foundations of thematic cartography which produces accepted compromises in data representation on paper, are now challenged by the capability of constructing spatial databases in which original field observations are stored and from which the interpretation and generalization models of experts can extract several specific themes. An example of this situation is the NATMAP initiative of the Geological Survey of Canada (Broome et al., 1993; GSC, 1990) which includes FIELDLOG, a relational database management system for data captured in the field (Brodaric, 1992; de Roo et al., 1993).

While the capability to store and retrieve at will many more observations (including their relations) than what can be placed as symbols on paper at a given scale, is a definite step towards the optimization of cartographic production, this new data representation allows 'time-resistant' and 'interpretation-explicit' data to be used for further decision-making and modelling processes. This means that new forms of thematic representation can be searched for, or that known types of representation can be improved upon or can be optimized, e.g., using statistical analysis.

Apart from general-purpose geological maps, the areas of mineral potential assessment and exploration, of geological hazard, and of environmental impact assessment are the primary targets for extended use of GISs. Typically, in those three areas, a first step in spatial representation is the construction of a database. A second step is the assessment of an existing situation, and a third step is the prediction of the situation in space and in time by means of models based on assessments in regions which are similar to the one under study. Invariably, a research component is fundamental until new standard representations will be solidified into acceptable cartographic products. This process is the unavoidable consequence of introducing GIS in institutes which have a public cartographic mandate, such as geological surveys, mines branches, departments of environment, etc. Similar considerations can be made for the introduction of 'processed' remotely sensed data into new experimental combined cartographic products where radar and other types of geophysical imagery (filtered and corrected by the elevation) is overlaid with geological maps. In this situation, the integration of different data layers has to be done by specialists who use their expert knowledge to make sure that the processed and combined 'information' layers sufficiently express the desired phenomena, and that the processing is explained and understood by the end user of the new thematic map products.

Because GIS technology is still fairly new and it is likely to have far reaching modifying effects on institutions and on individuals, it may be beneficial to summarize in the next Section some of the typical guidelines on GIS implementation.

## 6.2. HOW TO IMPLEMENT A GIS

The implementation of a GIS is a complex procedure that covers the analysis of needs of a laboratory, of a project, or of an institution, the development of awareness of GIS

capabilities, the setting up of a pilot project, the identification of software and hardware, down to system acquisition and training of staff. In Table 6.1 the views of three authors are listed for discussion. To some extent, the emphasis of the three sets of guidelines are on analysis (Burrough, 1986), on management (Aronoff, 1989), and on production (Clarke, 1991).

Burrough (1986) considered the sharpness in task definition as a fundamental aspect in the selection of a GIS. Vector and raster data structures are seen a complementary in its construction. Clear distinction in the skill of the personnel operating a GIS leads to the identification of the costs attached to the different tasks of data input/output, usage of analytical software and management. Aronoff (1989) stressed the need and the techniques to develop awareness of GIS technology within public and private institutions, to identify the most likely users, to initiate a pilot project which includes training of staff, database construction and system maintenance. Then, the critical issue of responsibility for services, maintenance and performance become fundamental. Clarke (1991) focuses, at an early stage, on cost-benefit analysis and a pilot study. This is followed by request for proposals, benchmarking, and cost-effectiveness evaluation which lead to the final implementation contract.

TABLE 6.1. THREE COMPLEMENTARY VIEWS ON SELECTION, IMPLEMENTATION AND ACQUISITION OF A GIS

| **Choosing a GIS (Burrough, 1986)** |
|---|
| **Definition of needs** |
| 1. Kind of user and user requirements: a. users with exact defined task (mapping and inventory agencies); b. part only of task defined exactly (environmental mapping agencies); and (c) no part of task work is defined exactly (university research and teaching groups, research institutions) |
| 2. Scope of application |
| 3. Technical choice – vector or raster? Complementary modes: vector mainly for archiving, network analysis and quality drawing, and raster for simulation and modelling? |
| 4. Finance available |
| 5. Personnel available: low-skilled and high-skilled |
| 6. Organizational aspects |
| 7. Cost of GIS computer and staffing requirements for: |
| 7.1. Data input; |
| 7.2. Data output; |
| 7.3. Use of data analysis software; and |
| 7.4. Planning and management. |
| **Procedures for setting up a GIS** |
| 8. Detailed inventory of existing work situation |
| 9. Anticipation of future data handling needs |
| 10. Kinds of hardware and software |
| 11. Looking for suppliers of software |
| 12. Considering hardware to support software |
| 13. Preparing a document detailing the requirements |
| 14. Benchmark test |
| 15. Choosing a system on its price and performance |
| 16. Plans for staffing and training needs |
| 17. Conducting a number of small test projects to train staff and to show managers what the system can do |

TABLE 6.1. (cont.)

**Implementing a GIS (Aronoff, 1989)**

**Phase 1**: Awareness of GIS technology and its benefits in organization

**Phase 2**: Developing system requirements and identification of users

**Phase 3**: Evaluating alternative systems (user friendliness and benchmarking)

**Phase 4**: System justification and development of implementation plan (database development, setting up of staff tasks, cost-benefit analysis)

**Phase 5**: System acquisition and start-up (maintenance contract, software training, database start-up, pilot project)

**Phase 6**: The operational system (initial database is complete and operating procedures have been developed, including maintenance, upgrade and services): who is responsible to provide services and guarantee performance?

**The GIS acquisition model (Clarke, 1991)**

**Stage 1**: Analysis of requirements
1. Definition of objectives
2. User requirement analysis
3. Preliminary design
4. Cost-benefit analysis
5. Pilot study

**Stage 2**: Specification of requirements
6. Final design
7. Request of proposals

**Stage 3**: Evaluation of alternatives
8. Short listing
9. Benchmark test
10. Cost-effectiveness evaluation

**Stage 4**: Implementation of system
11. Implementation plan
12. Contract
13. Acceptance testing
14. Implementation

The three approaches are clearly directed towards institutional GISs. For an individual researcher or a small research laboratory, while similar sequences of steps are likely to be required, the entire decisional process will have to be performed by a single person. This requires to identify the qualifications and the ideal background of the operator of the GIS and whether such experience is easily found in an individual specialist.

## 6.3. STAFFING QUALIFICATIONS: WHO OPERATES THE GIS?

In assessing the staff requirements for operating a GIS, we must clarify first the type of GIS activity (e.g., production oriented or for analysis and modelling), and the different levels of skill required for the various processing tasks. Burrough (1986) distinguished

'low-skilled' staff, those who are not supposed to know how a GIS works, but who must provide the input that make a GIS work and make sure that acceptable results are generated and 'high-skilled' staff, i.e. managers, liaison officers, technical experts and scientists. For each type of skilled personnel, costs for the different tasks (i.e., data input, data output, data analysis, planning and management) can be computed, which are proportioned to the degree of responsibility and expertise implied in the tasks.

The situation in the present discussion is mainly related to scientific research leading to advanced methods of GIS for information representation, assessment and prediction in the earth sciences. For this reason, the main qualifications for our GIS users are as follows:

**(a) Earth science background specific to the target application**

This background will be fundamental in understanding geological processes and in formulating models for GIS analysis.

**(b) Experience in photointerpretation, field work and thematic mapping**

The human vision system integrates images for interpretation. To understand the physical meaning of the features interpreted in images, a relationship must be established with what is observed in the field (ground truthing). Furthermore, what the computer will be working on is an interpreted set of digital features either digitized from photographs or coming from enhanced or classified images from which relevant information has been extracted, again by photointerpretation. The latter interpretation, enables the integration power of expert human vision to complete the results of extracting digital features which often appear disaggregated or disrupted.

**(c) Basic knowledge of statistical analysis**

Many aspects of the observed phenomena have a statistical signature in a GIS database, e.g., the frequency of a mineral or a structural association, therefore statistical analysis is needed in GIS model building.

**(d) Basic knowledge about microprocessors, their operating system and at least one programming language (BASIC, FORTRAN, Pascal, LISP, C or $C^{++}$)**

This knowledge will help in overcoming the vagaries of computer interaction, in programming interfaces and algorithms for tasks not planned in the available GIS software.

**(e) Experience in computerized data management**

Much GIS processing is management, therefore this experience facilitates the construction of analytical databases.

**(f) Experience in GIS and image processing techniques**

GIS and image processing are fields where datasets and data structures are frequently transformed. For these transformations, information is extracted, interpreted, and transmitted to modelling systems. Knowledge of the processing techniques enable to

evaluate and to anticipate the accuracy of the transformations and the validity of the processed digital data.

**(g) Willingness to communicate with experts in different disciplines which are relevant to specific applications or to acquire knowledge of those disciplines within the context of the target application**

**(h) Managerial experience including project planning**

Most of the activity and cost in the construction of a GIS project is in the creation of a database. Project planning is essential to ensure that, eventually, the analysis will have a chance of taking place!

Data integration requires extensive communication with experts in different disciplines.

These qualifications might not always be available in a single individual, however, they are desirable for the successful development of a GIS integration effort. They are not the qualifications of a 'superman' but they describe the background of the conventional GIS operators who are conducting research in resource exploration, in natural hazard studies, or in environmental impact assessment, three inseparable aspects of applied earth sciences. A multidisciplinary team effort is required in most GIS projects. This implies that, in general, operational knowledge should be preferred in different staff members rather than in a single individual. Training initiatives which aim at such multidisciplinary and cross-disciplinary backgrounds will be discussed in Section 6.6.

## 6.4. SOFTWARE

Evaluating and comparing software is a difficult task due to the variety of solutions to GIS problems. According to Clarke (1991, p. 477), for GIS evaluation "there is minimal public domain literature comparing the various systems, and their features are changing so rapidly that any such literature is soon out of date".

Several PC-based GIS software packages were listed in Table 2.5 mainly to point out that there are systems with different design and characteristics which may be preferable for some applications. PCs were chosen because they represent low-cost platforms of very wide distribution and therefore they are available anywhere, including developing countries. Here we want to identify types of software which, if not directly related to GISs, are of complementary use for programming, reporting and analysis. Figure 6.1 shows a network of contacts and of partial overlaps between different types of commonly available off-the-shelf software packages and GIS/RS-IPS. From the user's viewpoint, it is important to understand the relationships which exist between the various pieces of software.

Word processing programs enable text to be edited for documentation, where ASCII characters can be transferred back and forth the GIS and graphic output can be inserted in the text. Spreadsheet programs permit computations to be made on numerical tables that can be transferred as ASCII files. CAD programs, in addition to cartographic production, facilitate the composition of illustrations, annotations and can be used to generate databases with a linked number of attribute and of graphic files. Geostatistical and other contouring packages allow optimal interpolation processes and the transfer of interpolated data as raster datasets. Statistical analysis packages allow the analysis and transfer of

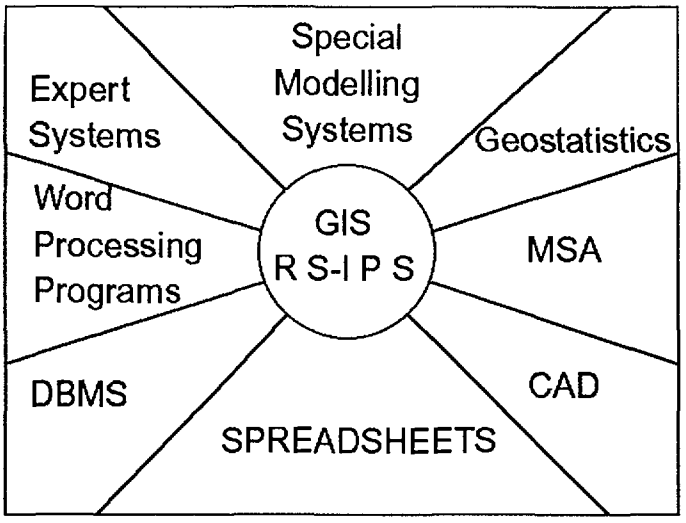



*FIG. 6.1. The relationships between GIS-remote sensing image processing software and complementary software for data analysis and management.*

multivariate tabular data optimizing regression and classification. DBMS software optimizes the processing and management of large attribute databases and the extraction of smaller subsets of data to simplify GIS processing. Expert shells, expert systems and special-purpose modelling programs, are used in conjunction with a GIS, in developing rules for modelling or in the fine tuning of assessment and prediction models.

What differentiates all these packages is their main focus. While any software can be seen as peripheral from the point of view of the other ones, it is important that a GIS performs over 90% of its tasks, leaving the less frequently and more specialized tasks to other specially optimized systems. This situation avoids overload on any one system with applications which might be of infrequent use for many users. Most analytical tasks in GIS processing will require interaction with those external software packages.

Interactive graphic environments such as Windows and Macintosh GUI (graphic user interface) enable the user to communicate largely by pointing at icons on the monitor, thus avoiding the use of the keyboard. At present, the difference between an operating system and a visual interactive interface has become rather subtle. For PCs, a new development of Windows, Windows-NT, will be made available in 1993 as a full operating system. Most programming languages are beginning to have visual versions, under Windows environments, to facilitate even further the interactive process (e.g., visual BASIC). Various tools are also available for individual self training. In the field of GIS and remote sensing, a global change electronic encyclopedia termed GEOSCOPE and RTUTOR, a radar remote sensing training tool (RTUTOR, 1992; Canadian Space Agency, 1992; Simard, 1992a, b) use Windows like techniques (hypermedia) to train users through the complexities of spatial data processing and of sensor characteristics and applications. This trend is likely to expand in the near future into special training packages and training centres.

## 6.5. HARDWARE

The development of hardware for PCs, and Macintosh platforms is in a continuous state of flux, so much so that anything that might be said at present is certainly obsolete by the

time it reaches the reader. Apart from this eventuality, we can see a consistent trend for PCs to remain the microprocessors with the widest distribution and accessibility. For research work in the field of GIS and of image processing of remotely sensed data, available and recommended hardware is listed in Table 6.2 (to hardly cover two years to come, say 1993–1995!).

TABLE 6.2. AVAILABLE AND RECOMMENDED PC-BASED HARDWARE FOR GIS AND IMAGE PROCESSING OF REMOTELY SENSED DATA

| HARDWARE | MINIMUM | RECOMMENDED |
|---|---|---|
| -CPU | 386 | 586 |
| -RAM | 8 Mb | 32 Mb |
| -math coprocessor | yes | yes |
| -processing speed | 33 MHz | 90 MHz |
| -memory cache | none | 256K |
| -printer | laser | laser |
| -plotter | A3 inkjet | A0 inkjet |
| -digitizer | A3 | A0 |
| -scanner | none | A1 |
| -camera | none | $512^2$ chip |
| -mass storage | 500 Mb | > 1 Gb |
| -graph. disp. cards | SuperVGA (8 bits: $2^8$ colours) | SuperVGA (24 Bits:$2^{24}$ colours) |
| -backup | tape streamer Bernoulli Box | optical disk |
| -CD ROM | none | yes |
| -CD-R (recorder) | none | yes |
| -communication (LAN) | none | yes |

PCs are essentially single user systems and may not be the desirable platforms for running large jobs or for processing large arrays of data. They have become faster machines, however, and often their response time can be shorter than what can be obtained on a workstation with two or more simultaneous users. The lower price for PCs has resulted in the development of a vast amount of software for a worldwide market. For this reason, and because workstations are more expensive and less available, numerous software developers are still hesitating to direct their efforts towards workstations. Many low-cost storage media for PCs now exist, such as CD-ROMs for which the price is little more than that for a hard disk drive. CD-Rs (or CD recordable) drives are now available for less that US $10 000. They considerably expand the capabilities of PCs and are now up to the requirements of an image processing system which generates one or more images in most processing steps. Local area networks or LANs, enhance even more the power of microprocessors by enabling several CPUs to share large storage media. New graphic cards, even if not yet standard products, are making high resolution colour

visualization very accessible, e.g. superVGA or mega-super-VGA. Today, a critical problem is data integration and analysis, not any more the limitations of hardware: the complexity of formulating a real problem in a computational manner.

## 6.6. TRAINING NEEDS AND INITIATIVES

The need to develop applications of GISs for global environmental problems has encouraged several initiatives of the United Nations. In 1985 the United Nations Environment Programme (UNEP) has implemented GRID, the Global Resource Information Database, to provide scientists and planners around the world access to an integrated GIS database (GRID, 1991; UNEP, 1992). Countries like Kenya, Switzerland, Norway, Thailand, Japan, Nepal, Poland, the Russian Federation, Fiji, Brazil and Canada are supporting the GRID system either in research and development or in technology transfer. Training in the use of GISs and remote sensing technology has been undertaken through UNITAR, the UN Institute for Training and Research.

The International Institute for Aerospace Survey and Earth Sciences (ITC), in the Netherlands, has specialized in graduate training in GIS, remote sensing and image processing in areas of mineral exploration, geological hazards, landuse planning, urban geography, cartography and environmental management. Several training packages are being made available by ITC, such as GISSZ, a landslide hazard assessment database and GIS training software (van Westen, 1993).

A core curriculum study for a 9-month academic training in GIS was made available in 1989 by the National Centre for Geographic Information and Analysis, NCGIA (1989), a centre of excellence consisting of three universities in the USA (University of Main in Orono, University of New York in Buffalo, and University of California in Santa Barbara). Several universities in North America and Europe now have intensive training in GIS and remote sensing. Truly interdisciplinary training, however, is still a long way to become a reality due to the difficulties to transform a monodisciplinary culture into a multidisciplinary one.

## REFERENCES

Brodaric, B., 1992, Map compilation with CAD for geological field mapping. Proc. Computers and Mineral Exploration (CAME) Seminars, Toronto, Canada, March 28, 1992, 13 pages (unpublished manuscript).

Broome, J., Brodaric, B., Viljoen, D., and Baril, D., 1993, The NATMAP digital geoscience data management system. Canadian Journal of Earth Sciences, (in press).

Canadian Space Agency, 1992, GEOSCOPE, global change encyclopedia. CSA, External Relations, 500 Rene-Levesque Blvd. West, Montreal, Quebec, Canada H2Z 1Z7.

GRID, 1991, GRID News, v. 4, n. 4, P.O. Box 30552, Nairobi, Kenya, 12 p.

GSC, 1990, NATMAP Canada's national geoscience mapping program. Geological Survey of Canada Open File 2256, Report of Workshop held in Ottawa during March 8-10, 1990. St. Onge, Editor, 83 pp.

IDRISI, 1992, IDRISI News: Version 4.0, TOPOS and UNITAR GIS Workbook Series Spring 1992, v. 4, n. 1, 6 pp.

NCGIA, 1989, Introduction to GIS: NCGIA Core Curriculum, Draft Version, University of California in Santa Barbara, Eds Goodchild M.F and Kemp K.K., July 1989, 3 volumes, 75 lectures.

de Roo, J.A., van Staal, C.R., and Brodaric, B., 1993, Application of FIELDLOG software to structural analysis in Bathurst mining camp, New Brunswick. In, Current Research, Part D: Geological Survey of Canada, Paper 93-ID, pp. 83–92.

RTUTOR, 1992, The Radarsat Tutor: Version 1.0, Rev. B. Vantage Point International, 1324 Cedarcroft Crescent, Gloucester, Ontario, Canada K1B 5C8, and Brilliant Eyes, 24 Davidson Street, Kingston, Ontario, Canada K7L 5G5, 15 p. and 4 diskettes.

Simard , R., 1992a, A global change encyclopedia. Proc. of The Canadian Conference on GIS-92, March 24-26, 1992, Ottawa, Canada, pp. 949–960.

Simard, R., 1992b, Geoscope: an interactive global change encyclopedia. Proc. of Meeting for the International Space Year, March 1992, Munich, Germany, 5 pp.

UNEP, 1992, GRID Summary Data Bulletin Available Datasets (Nairobi). Compiled by GRID-Nairobi, August 1992, 63 pp.

van Westen, C.J., 1993, Application of Geographic Information Systems to Landslide Hazard Zonation. Enschede, Netherlands, ITC Publication 15, 245 p. (Technical University of Delft/ITC, Ph. D. Thesis).

# 7. CASE STUDIES ON DATA INTEGRATION FOR MINERAL EXPLORATION

## 7.1. DATA INTEGRATION AND GIS MODELLING IN THE NORTHWEST TERRITORIES OF CANADA

E.M. Schetselaar, A.G. Fabbri
ITC, Netherlands

### 7.1.1. Introduction

A case study on data integration and GIS modelling was constructed from a preliminary study by Fabbri et al. (1989) and Fabbri and Kushigbor (1989) in the Bathurst Inlet area of the Northwest Territories of Canada. Reconnaissance fieldwork and regional compilation, including a mineral potential assessment of the area was done by Roscoe (1984). Detailed fieldwork (1:20 000–1:10 000) by the Geological Survey of Canada, between 1988 and 1991, took advantage of several enhanced plots of SPOT, airborne SAR images, and TM colour composites. Field notes for that fieldwork (Jefferson et al., 1990) and a detailed multiple data set for part of the study area were used to develop a realistic case study to train exploration geologists. The practical approach to such integration was based on hands-on work using PC-based tools for:

(1) capturing field data using a relational database, FIELDLOG©;

(2) manipulating cartographic symbols using a computer aided design and drafting system, AutoCAD©; and

(3) processing and modelling information using a hybrid GIS-Image Processing system, ILWIS©.

The approach aimed at the integration of primary field data with remotely sensed data, and with other secondary data digitized from maps. Integration is seen as a modelling process of decision-making which can lead to new cartographic techniques where mapping and modelling are inseparable tasks for information and knowledge representation (leading, for instance, to a map as a knowledge base stored on a magnetic diskette).

While the techniques of field data capture by a portable computer, and the usage of a CAD system for cartographic production in geology are beyond the scope of this case study, an overview of the integration and analysis part of that work is considered a useful topic in this chapter.

A brief description of the study area and of the data set provide the framework to the following analytical parts:

(a) processing of remotely sensed data;
(b) processing of geophysical data;
(c) performing data integration; and
(d) developing spatial analysis.

---

AutoCAD© - 1982-90,1992 Autodesk Inc. 2320 Marinshipway Sausalito, CA 94965.
FIELDLOG© - 1992 Geological Survey of Canada, Ottawa, Canada.
ILWIS© 1992 - ITC - International Institute of Aerospace Surveys and Earth Sciences, Netherlands.

This case study is extracted from Schetselaar et al. (1993) who developed a workshop tutorial out of their direct experience both in the field and in the classroom. The following is their view of data integration (Op. cit., p. 5):

"The increasing amount of multiple geological data sets readily available to geologists, has created a need for efficient capture, storage, management, retrieval and analysis of geoscientific data. Today the earth scientist is faced with the difficult task of relating and integrating vast amounts of data types of different nature, obtained from different sources and compiled at different scales. In order to use all these data for mapping, interpretation and modelling, the geologist should ideally be able to spatially link field observations with ancillary spatial data such as airphotos, gridded geophysics, satellite imagery and topographic data, and other point data, such as laboratory results from geochronology, geochemistry and thin sections. The tremendous growth in the development of GIS technology in the last decade has provided the geological community with a toolkit for addressing this data integration problem. The digital data format allows one to easily and efficiently output, update, compare, and visualize multiple data sets in a GIS. Moreover a GIS provides a wide range of spatial analysis capabilities for selective retrieval, statistical analysis and modelling of both spatial and non-spatial data. GIS is complementary to the geologist's inherent ability to think about the object of study in a spatial and integrated manner."

### 7.1.2. Description of the study area (geology, geomorphology, ecology)

Information about the geology, geomorphology, vegetation, soils and wildlife from the study area can be found in Zoltai et al. (1980), Roscoe (1984) and Jefferson et al. (1990).

Bathurst Inlet occurs along a major fault zone that runs along the western side of the inlet. The total displacement along this fault was about 140 km, with vertical movement of about 1 km. This created a deep basin, now drowned by the sea. Intrusive rocks that are resistant to erosion protected the underlying softer sedimentary rocks. This resulted in butte and cuesta topography in which vertical scarps may be as high as 300 m. The variety in relief near the Inlet is in sharp contrast with the surrounding area typical of the Canadian Shield, which is characterized by low hills and broad valleys that provide the only variation in relief.

A generalized geological map is shown in Fig. 7.1. A summary of the geology of the entire Bathurst Inlet area followed by a description of the geology of the Pistol Lake Area based on the report of Zoltai et al. (1980), Roscoe (1984) and a 1:10 000 geological map produced by C.W. Jefferson of the Geological Survey of Canada from which Fig. 7.2 was generated.

**Archean rocks.** Archean volcanic (mainly basalts) and sedimentary rocks (volcanoclastics, carbonaceous pyritic slate, chert, carbonates, lenses of banded iron formation), are believed to have been deposited from 2.7 billion years ago upon 2.9–3.2 Ga old granitic and tonalitic basement rocks. Granitic rocks were intruded by pre-, syn- and post-orogenic granitic rocks. The rocks were metamorphosed to greenschist, lower amphibolite and upper amphibolite facies. The volcanic and sedimentary supra-crustal rocks were tightly folded into the present near vertical attitude through several phases of intense deformation. The depositional setting for the Archean supra-crustal rocks is difficult to ascertain due to the poor structural and stratigraphic control in the area. In general, however, these rocks formed as a consequence of the

stretching and rupturing of sialic crust resulting in down dropping of a series of crustal blocks to form fault bounded sea floor basins. These basins grew through the extrusion of lavas on the sea floor. Turbiditic sediments were deposited over more extensive areas and in part record the active volcanism that must have persisted in this area during sedimentation. Dikes, sills and plugs in the area record the paths of magma that did not reach the surface. These Archean rocks are the focus of the tutorial study.

**Aphebian rocks.** The metamorphic and intrusive events culminated in the development of the Slave Structural Province (2.6 Ga ago). This Cratonic region was then uplifted and deeply eroded. Around 2.2 Ga, it was intruded by alkaline and peralkaline bodies,

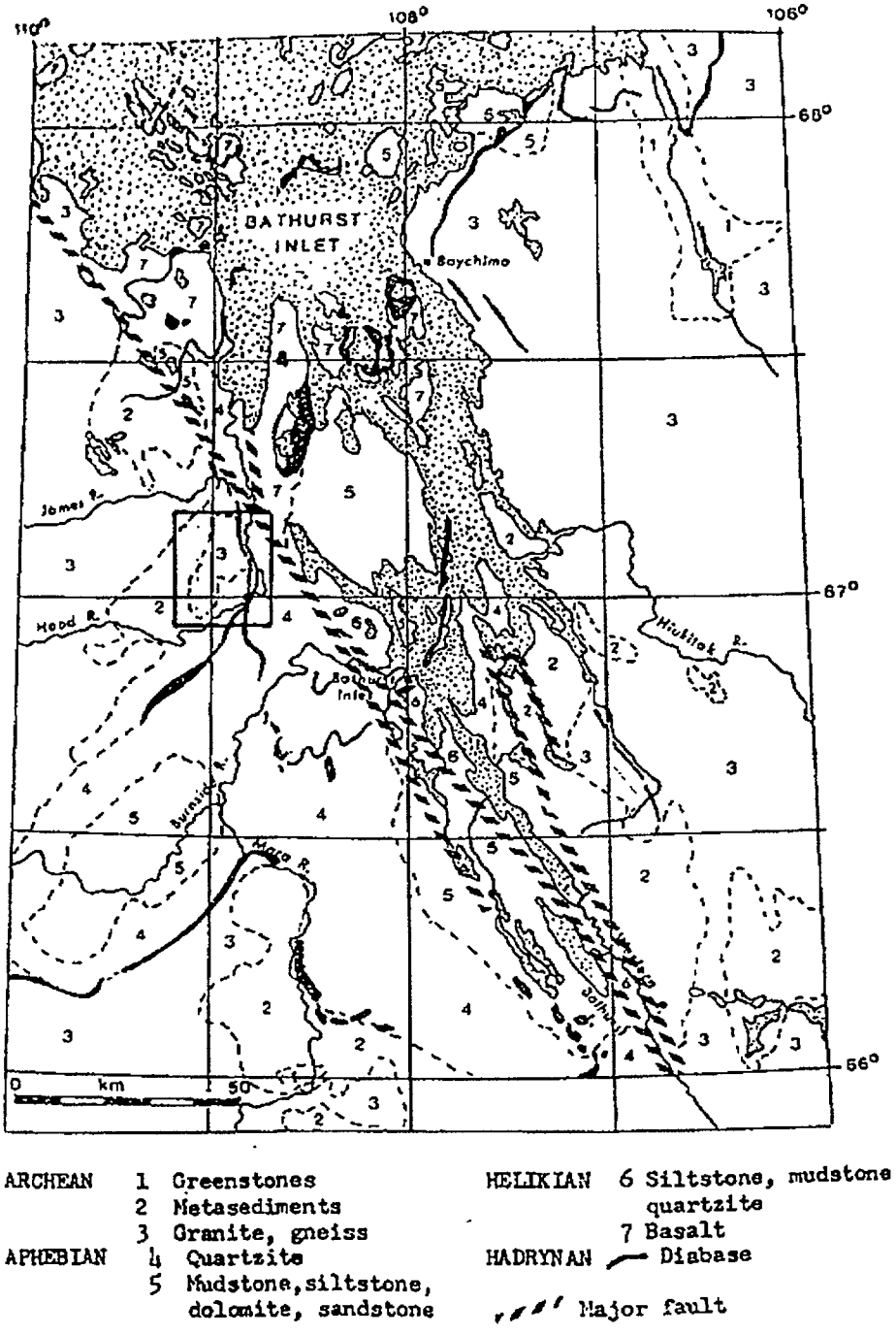



*FIG. 7.1. Generalized geological map from Bathurst Inlet.*

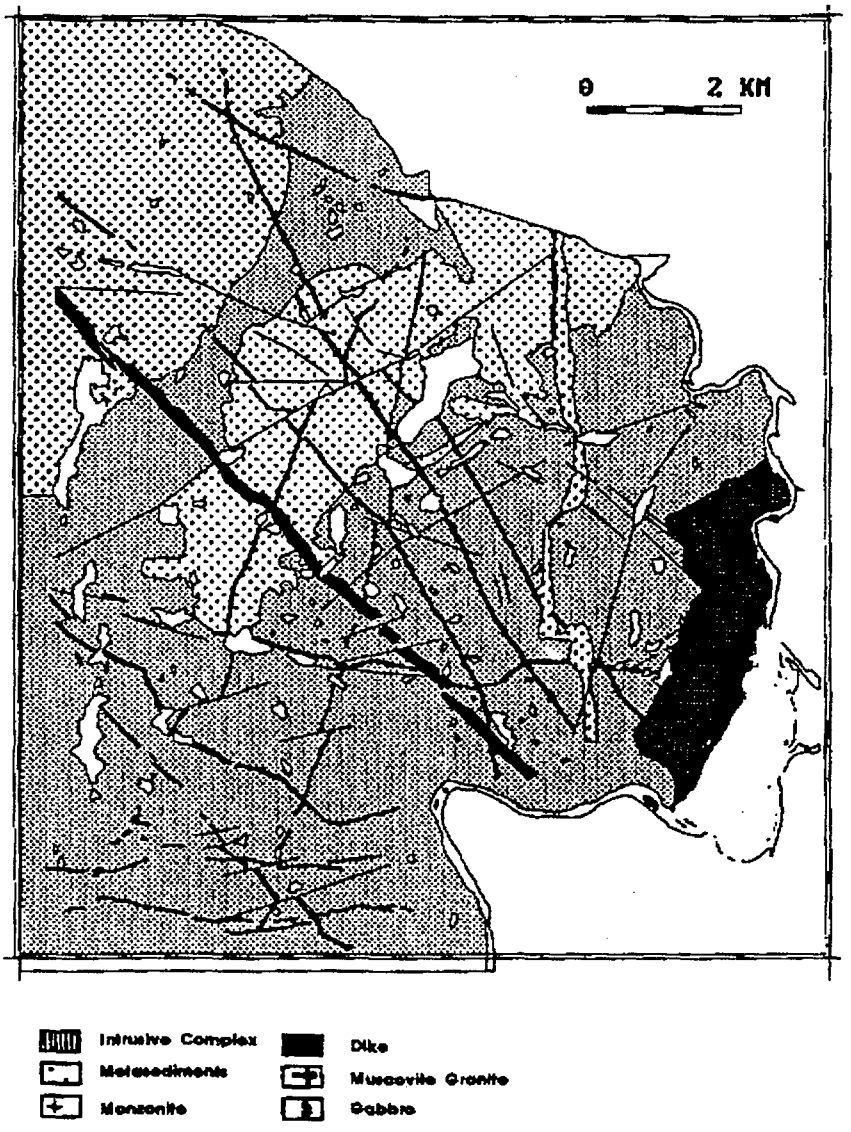



*FIG. 7.2. Geology map of the Pistol Lake Area, Hood River Belt, compiled and digitized from 1: 10 000 geology maps GSC (Henderson et al. 1991).*

probably along extensional fractures. Around 2.0 Ga ago, the Slave Province was subjected to tensional strain resulting in rifting with related warping and development of the northerly trending seaway. At Bathurst Inlet the Kilohigok intracratonic basin was formed. Sediments of shallow marine, deltaic and fluvial origin were deposited in this basin giving rise to a sequence of a great variety of sedimentary rocks (turbidites, stromatolitic and clastic carbonates, quartzites, argillites, lithic and arkosic sandstones, conglomerates and breccias).

**Helikian rocks.** Sediments of Helikian age occur in a succession of six formations and several members. Three phases of sedimentation, separated by unconformities were recognized each marking a major change in the depositional environment. The oldest formation consisting of arkoses, silt stone, conglomerates and breccias were deposited in a fluvial environment. The next sedimentary rocks consist of an alternation of carbonates, quartzites and conglomerates. The Ekalulia formation, overlying these older sedimentary rocks consists of lava flows of basalt. The pillowed and massive flows are overlain by the

Algak formation consisting of cross-bedded reddish arkose and siltstone. The youngest Precambrian rocks are sills and dikes of Hadrynian age. The sills, up to 100 m thick, consist of sheets of diabase intruded sub-horizontally into sedimentary beds. Dikes (up to several hundreds meters wide and several kilometres long, dip nearly vertical and consist mainly of diabase, but also may be diorite, gabbro and pyroxenite. These intrusive rocks and their older equivalents have a pronounced expression on the aeromagnetic maps.

**Bedrock geology of Pistol Lake Area.** The Pistol Lake Area is part of the Hood river belt mapped by Henderson et al. (1991). As can be seen from Fig. 7.1, the bedrock geology of the Pistol Lake area mainly comprises the Archean units of the Precambrian lithologies. Belts of meta-sedimentary rocks generally of younger age towards the East consisting of greywackes, amphibolites, phyllites and schists, alternate with granitic rocks, the latter subdivided into gneissic and massive granites. An unconformity along the East side of the area separates the Archean rocks from the Aphebian rocks. These Lower Proterozoic lithologies consist of various sedimentary rocks deposited in shallow marine, deltaic, and fluvial settings.

The Archean rocks between Pistol Lake and Hood River are intruded by a sheet like body of gabbro. To the South this intrusion is mainly emplaced along the unconformity between Archean rocks and the Western river formation. They are considered to be of Aphebian age. The youngest Precambrian rocks in the Pistol Lake area are diabase dykes. At least three and possibly four diabase dyke swarms exist. The oldest are East-West trending diabase dykes assigned a Late Archean age of 2692 Ma (Rb-Sr mineral and whole rock isochrone. The Southeast-Northwest trending dykes are part of the Proterozoic Mackenzie Dyke Swarm (with a baddeleyite U-Pb age of 1270 Ma). The youngest dyke is the northerly trending diabase dyke which intrudes the East side of the Pistol Lake area. In Fig. 7.2 the geological bedrock map is presented, which was originally compiled by the GSC on 1:10 000 scale by R. Wyllie (Henderson et al., 1991).

**Mineralization.** Gold is found in iron formation beds within the upper sequence of a turbidite unit (metasedimentary unit). The iron formations are mixed silicate, oxide and sulfide facies with gold commonly found in the latter. Available information suggests that the gold is not vein related and may be syngenetic or early epigenitic (Henderson, 1991).

**Quaternary geology.** The Bathurst Inlet area was glaciated during the Wisconsian stage by ice advancing from the Southeast. Upon deglaciation the depressed land was inundated by the sea. The following glacial and marine deposits are present in the Pistol Lake area (see also Fig. 7.2).

**Ground moraine.** Till deposits of uneven thickness blanket the terrain consisting of unsorted debris. In general the till is bouldery and stony, having a silty sand to sandy loam texture (Zoltai et al., 1980). The chemical properties of the till are variable, depending on the mineral composition of the source bedrock material.

**Marine sediments.** Marine sediments are fine grained (silts and clays) and occur on the Northeast edge of the study area in the Hood river valley. Chemical composition is slightly acid with a relative high nutrient levels (higher than basic and acid till deposits). This fact and the texture of the sediments explains the lush vegetation cover on these deposits (Zoltai et al., 1980).

**Fluvial deposits.** They are confined to the Hood river valley. The material is bedded sand and gravel. The chemical composition resembles the till of the source area.

**Vegetation.** Broad vegetation types were mapped by Zoltai et al. (1980) from field work and airphoto interpretation. The following is a description of the vegetation classes which occur in the Pistol Lake Area.

1. **Rock lichen type:** This vegetation type is found on bare bedrock, and on boulders. A variety of lichen species grow on the rock substrate which influences the composition of the community. Vascular plants occur in cracks or between stones. Also mats of a moss species (Rhacomitrium lanuginosum) cover large patches of bedrock.

2. **Dwarf shrub-heath type:** This vegetation type occurs on well to excessively drained till uplands. It consists chiefly of ground hugging dwarf shrubs. Also mosses may occur.

3. **Low shrub-heath type:** This vegetation type occurs mainly on well to imperfectly drained slopes on till materials. It consists of low willows and birch and heath vegetation.

4. **High shrub type:** This vegetation type is characterized by shrubs 30–200 cm high. Two subtypes were recognized, one occurring on lower slopes, imperfectly to poorly drained margins of ponds, and along creeks. The other type is common on imperfectly drained deltaic or fluvial sands.

5. **Sedge-Cottongrass Meadow:** This vegetation type is characteristic of imperfectly drained clay deposits. It is dominated by sedges but also mosses are common.

**Ecological land classification.** Ecological land classification, based on the Canada Committee of Ecological Land Classification was completed by Zoltai et al. (1980) in order to evaluate the elements of the physical environment in terms of their importance to the biological components.

Three levels of subdivisions have been made: Ecoregions, Ecodistricts and Ecosections. The entire area is part of the Low Arctic Ecoregion. The climate in this region is too severe for trees, but not severe enough to prevent the development of continuous vegetation. From the further subdivision it was derived that The Pistol Lake Area is part of the Bear-Slave as well as of the Bathurst Hills ecodistrict. The Bear-Slave ecodistrict is characterized by bedrock controlled relief and has the physiography of an elevated peneplain (average altitude is 400 m but local variations are less than 30 meters). Much bedrock exposure is present in the area but also deep till deposits occur. Large esker systems extend in Southeast Northwest directions and many lakes are present. The Bathurst Hills ecodistrict is characterized by a topography accentuated by steep cliffs with flat or gently sloping summits. Spectacular waterfalls occur on all major rivers. Bedrock outcrops are common, alternating with thick till deposits. Marine sediments can be found in broad plains or in isolated pockets among rock ridges. Ecosections were mapped on the basis of: (1) broad relief, (2) texture, and (3) petrography of soil materials and their thickness. In most of the area, there are terrains with low to moderate relief bedrock exposures and till deposits with both acidic and basic compositions are prevailing. On the northeastern and southeastern parts of the area, however, sediment cover of fluvial and marine origin, supporting lush vegetation growth, is predominant.

### 7.1.3. The data set

A multiple data set was constructed to provide exploration geologists with the context for interpreting and exploiting all ancillary data of potential utility in mapping. The data set is a typical one which is often available but with requires some interdisciplinary effort to be properly used. Most geologists are traditionally taking little advantage of such data because of the complexity of integrating the data in a systematic manner within a GIS. In general, Lack of training and the often overwhelming pressure of tradition leave little room for innovative techniques to be employed. The following is a description of the work files for the workshop tutorial after Schetselaar et al. (1993).

**(a) Hard copy imagery**

- Colour copy airphoto comprising the area directly north of the Hood river with geological field stations and photointerpretation on overlays.

- Landsat Thematic Mapper colour composite image bands 7, 5 and 3 from colour copy photoprint of filmwriter diapositive.

- Colour enhancement of part of the Pistol Lake area, fusing aeromagnetics and TM band 4 (edge enhanced) by IHS transformation.

**(b) Digital data**

- Landsat TM imagery, bands: 1, 2, 3, 4, 5, 6 and 7 provided by the Canada Centre for Remote Sensing ,CCRS, in Ottawa;

- SPOT Panchromatic band, also provided by CCRS (see Fig. 7.4, Plate 12);

- Digital elevation contours digitized from a 1:25 000 topographic map (see Fig. 7.5, Plate 13);

- High resolution aeromagnetics flown by Silverheart Mines, in 1989 (see Fig. 7.6 (b), Plate 13);

- Geology coverage digitized from 1:10 000 geology map (Henderson et al., 1991; see Figs 7.2 and 7.6 (a), Plate 13);

- Tables containing structural data obtained from the geology map mentioned above;

- Banded iron formation outcrops digitized from the geology map;

- Faults and shear zones digitized from the geology map;

- Locations of known mineral occurrences digitized from a GSC open file publication by Roscoe (1984).

Several colour illustrations are a visual account of this data set. Figure 7.3 (Plate 11) shows Landsat TM colour composites of the study area. Figure 7.4 (Plate 12) shows part of a SPOT panchromatic image and a scanned aerial photograph. Figure 7.5 (Plate 13) shows a digital elevation image and Figure 7.6 (Plate 13) shows part of the geological

map of the Pistol Lake area and the corresponding aeromagnetic image. The data set realistically represents a situation in which raw data have to be analyzed, enhanced, modified, and combined, in order to extract the information to use in exploration modelling. Contrary to what it may be expected by the uninitiated, only a small part of a data set will eventually be usable for such modelling. Much expert knowledge and awareness of the information content of the database are required to finally translate observations into a computational form for GIS modelling. This data set offers the opportunity for an introduction to geoscience data integration.

### 7.1.4. The exercises

These exercises have a simple introductory value for geologists. The strategy of the tutorial is of explaining the background of the data and of the analyses while guiding in the interaction with a computer. The tutorial exercises are an introduction to the more advanced work to follow in Sections 7.2 and 7.3 where real life applications of an individual researcher and of a geological survey team are described.

**Processing of remotely sensed data**

Important in the interpretation of remotely sensed data from humid and moderately to densely vegetated environments, such as the tundra, is to gain understanding of the spectral responses of the various ground cover classes (vegetation, water, bare soil, bedrock, etc.) that mask spectral responses from pure bedrock. This is not only crucial information, to be able to efficiently exclude the pixels of the scene that mask the spectral response of the bed rock, but the patterns in these ground cover classes might also be useful for geological interpretation. Such patterns are, at least partly, controlled by bedrock geology. Geomorphology, landscape ecology and geobotany are the fields that study these relationships. For this reason background information on physiography, sediment cover (Quaternary geology) and vegetation is provided in the previous section on the study area.

The exercises start with the generation of several displays of Landsat TM bands and of their grey level histograms. The displays are useful to become familiar with the variety of digital number values and consequently to relate them to the differentiation of some geological units on account of their spectral responses. Contrast enhancement is performed next, to exclude the pixels corresponding to water which have low values. The visual interpretation of the bands is improved.

Colour compositing follows of sets of three bands and the different colours of the composites are related to the physiography of the landscape. Figure 7.3 (Plate 11) shows two colour composites for the Landsat TM image of the study area. Statistical criteria for the selection of optimal combinations of three bands are used which relate the sums of pairwise band correlations with the sums of standard deviations of the grey levels of each band in the combination. In addition, a measure of spectral contrast between band pairs is obtained by the biomass or green vegetation index. Vegetation, water, and fluvial deposits are masked to isolate bedrock exposures. An image is obtained which contains pixels for which the spectral responses are mainly controlled by bedrock, however, they probably are still influenced by lichens and moss vegetation communities.

To be able to fit remotely sensed data in overlay with maps at a later stage, the Landsat TM imagery has been geometrically corrected in respect to a topographic reference. A

SPOT image has to be registered with the TM image. Figure 7.4 (a) (Plate 12) shows the SPOT panchromatic image of part of the Pistol lake area.

**Processing of geophysical data**

Potential field surveys are carried out in order to map local variations in the earth's main fields (gravity and magnetic). The variations may then be interpreted as due to physical property distributions (density and magnetism) in the rocks of the earth's crust. Aeromagnetic surveys are relatively inexpensive and widely available. For this reason they are the most utilized geophysical data in mineral exploration surveys.

In recent years qualitative interpretation of aeromagnetic data has improved significantly by the use of image processing systems by presenting the aeromagnetic data as raster images. Each grid cell has a value assigned to it corresponding to the geophysical variable: aeromagnetics or gravity. Compared to the interpretation of conventional contour maps raster representations can be easily interpreted and digitally processed.

Aeromagnetic data is inhomogeneously distributed (high density along flight lines and low density perpendicular to the flight lines). A consequence of this is that gridding (interpolation) procedures may lead to serious degradation of the data. Sophisticated interpolation procedures are required to compensate for the inhomogeneous measurement distribution. By choosing different parameters, it is possible to find a compromise between the two evils: not loosing to much information along flight lines and not creating artificial patterns (bulls eyes) due to the inhomogeneous data distribution.

In the exercises on geophysical images, the raster representation of the high resolution aeromagnetic data is displayed similarly to the Landsat TM bands, however, contrast enhancement techniques can be used to highlight subtle variations in specific ranges of values, for instance using different histogram stretches, pseudo-colour look-up tables, or defining optimal transfer functions for colour assignment. Artificial illumination or hill shading by filtering is performed on the high resolution aeromagnetic raster image and on the image of topographic elevation. Contrast enhancement is then applied to the hill shaded images.

To gain insight in the geology in a certain area we often want to analyze one data type in conjunction with another. For example we might want to verify if a geophysical anomaly is related to structural features observed at the surface. Figure 7.5 (Plate 13) shows part of the geological map of the Pistol Lake area and the corresponding high resolution aeromagnetic image. The coincidence of subtle patterns in different digital images may lead to more confidence in the interpretation of those features. On the other hand patterns present in only one or the other image might supplement the features that can be extracted from the combined interpretation.

To integrate digital images of different sorts, we can not use the red, green and blue colour representation because the numerical characteristics are not represented by uniform colour gradations and are therefore difficult to interpret. We have to transform the cartesian colour representation into quantifiable colour attributes that can be distinctly perceived. Such a colour coordinate system is the IHS transform in which: I = intensity, i.e., total brightness of a colour, H = hue, i.e., average wavelength of a colour and S = saturation, i.e., purity of a colour.

To integrate the aeromagnetic data with the geometrically corrected SPOT panchromatic image the procedure may be as follows. The variations in gammas of the geophysics are assigned to the hue, and the digital numbers of the SPOT image to the intensity, while keeping the saturation constant. In image processing systems, such an approach involves an IHS to RGB transformation in which 3 image bands are displayed on a total of 24 bits displays (8 bits for RED, 8 bits for GREEN and 8 bits for BLUE). An alternative approach is however required to do such a transformation on a 8 bit display device such as the commonly available VGA (or superVGA) graphic card which is limited to the display of a total of 256 colours at one time. The method consists of slicing the geophysical image in 15 grey levels, the SPOT image into 16 grey levels, and then of combining them.

**Data integration**

Having performed several enhancements on a multiple data set, it becomes useful to compare our field observations with the patterns visible in the enhanced raster images. Each enhanced product can contribute to a different aspect of a final geological model. For example, geophysics could give us extra information in those areas where we have no or sparse outcrop. Enhanced Landsat imagery and the hill shaded elevation image could be useful to extent observed lithological contacts, faults and shear zones. In addition, the field observations stored as point data can be displayed on the images in an integrated fashion, while providing ground truthing to our interpretations.

The combined use of field data and remotely sensed data, if used properly, can bring more information to the interpreter than can the two data sets on their own. The simple techniques used here are applicable in other fields as well, were ground based geoscientific information is important. These techniques allow the user to think about the spatial modelling which will be discussed for the last exercises.

By using photo interpretation techniques on the different raster images, and overlaying field observations on them, a final map of the geology can be compiled. The georeferenced and enhanced raster images are displayed as backdrop images with an overlay of the observed lithologic contacts and structural readings. With the cursor of the digitizing tablet lines can be traced on the image eventually having a topographic map referenced on the tablet.

This process is termed 'on screen digitizing.' A final interpretation is made by digitizing on the background of the available enhanced raster images to complement the geology vector map already available. The interpretation is extended to the observed lithologic boundaries and use is made of the structural readings for interpolation. The interpretation is traced on the currently loaded background image or other enhanced products are loaded as raster background images. Also a lineament interpretation is made and bedding/cleavage relationships are used to infer mayor fold axial traces.

The geological map was initially digitized in the vector data structure. The vector representation implies that a topology was defined which describes the spatial relationships between coordinates, lines or segments (consisting of an array of coordinates) and polygons (the areas enclosed by the segments). In the raster data structure the map consists of a regular array of square grid cells in which each cell corresponds to a unit area on the ground. The grid cell or pixel value corresponds to a certain lithologic unit.

In general, the advantages of using the vector data structure are that it requires much less storage space, data can be efficiently rescaled or transformed to other cartographic projections and the links with attribute data are retained even after having performed several spatial transformations. However, if we want to use the geology in spatial analysis, that is if we want to combine the geological coverage with other more readily available georeferenced raster data, then a better structure is the raster data structure. For example, geological modelling often is undertaken by manually overlaying gridded geophysical, geochemical or other point data, remotely sensed data, scanned air photographs, etc. Modelling in the GIS environment allows facilitating this interpretive process that is often intuitive to the geologist. This has the advantage of being much more efficient, quantitative and versatile.

This exercise can be considered an introduction to raster modelling by the use of a raster processing techniques. The geology is rasterized from a polygon map using as a reference the geometrically corrected Landsat image. This implies that the geology is rasterized on the georeference of the Landsat TM image (a pixel size of 30 meters, and the coordinates assigned to the pixels are of UTM zone 12). Pixel values are assigned according to the alphabetic order of the lithology names: the dykes will be attributed a pixel value of 1, granite will have the value of 2, etc. To see this relation, the raster information file is displayed. With the geology map in raster format it is now possible to perform various kinds of spatial analysis operations. One useful application in this study is to compare the geological map compiled from the observed boundaries and the interpretation of the image displays with the original geology map compiled by the Geological Survey of Canada (Henderson et al., 1991; Jefferson et al., 1991).

To assess the correlation between the two maps, all occurring combinations of pixel values in both raster maps need to be identified. It is necessary to evaluate to what extent the on screen interpretation overlaps with the existing geology map. An operation termed 'cross' finds all occurring combinations of pixel values corresponding to lithologic units in both raster maps. Its results are stored in a cross table which contains all unique combinations. To perform the analysis properly, one has to make sure that both maps are encoded consistently, i.e., both maps should have the same range of pixel values and the pixel values should be assigned to the same units.

**Spatial analysis**

In the following set of exercises spatial analysis is performed to define/extract areas with a high probability for gold mineralizations. It is done on a qualitative basis, and with a hypothetical model, mainly because not enough information is available about styles of mineralization in the study area and because there are not enough known occurrences of gold in it to do statistical modelling. Examples in which statistical modelling could be applied to quantify probabilities for mineralizations were discussed by Agterberg (1989) and by Bonham-Carter (1990).

Let us suppose that the Vice President of Exploration of a mining company working in this area and is trying to test a hypotheses for gold mineralization. Field information is extremely limited, but it is known that gold mineralization in the area is often related to banded iron formation. Firstly, the mapped units of banded iron formation and their neighbourhoods are considered as targets for exploration. Secondly, the shear zones that intersect the banded iron formation are treated as higher priority targets for exploration.

The georeferenced data set on which the hypotheses are going to be tested consist of the following:

*Vector maps:*

| | |
|---|---|
| **bif** | Digitized segments of the banded iron formation lenses |
| **fault** | Digitized lineaments from previous exercises |

*Raster maps:* (pixel size of 10 meters)

| | |
|---|---|
| **geolw** | Window from raster map geology covering the eastern part of the Pistol Lake Area |
| **bif** | Raster map of the banded iron formation lenses |
| **prospect** | Raster map of the known gold occurrences |

*Table:*

| | |
|---|---|
| **prospect** | Table containing attribute information of the known gold occurrences |

**Retrieval of attribute information on mineralizations**

In the Pistol lake area there are a few locations with known mineral occurrences (Roscoe 1984). Attribute data of these occurrences are stored in a table. In a GIS, the capability is provided to link attribute data with their location on a map. This linkage can be established in many different ways. It is important that the GIS users understand and be able to implement these linkages so that attribute information can maintain its spatial integrity. Not knowing where mineralization occurs in real space it makes it impossible for the explorationist to perform spatial modelling. In this example the attribute-spatial linkage was established when the point locations of the occurrences described in the table **prospect** were rasterized and the map **prospect** was created.

Attribute information such as grade, width of mineralized zone, etc., from those points is retrieved. The map of known gold occurrences is displayed on the colour monitor. A zoom window with a cursor is displayed on the colour monitor and while the cursor is moved to the known occurrences, the attribute information is also displayed to become familiar with the database stored in the table.

**Overlaying segments on a raster background map**

To check visually the spatial relationship between the known deposits and the gold occurrences the banded iron formation lenses are displayed on top of the map of known gold occurrences. Then, the segments of the BIF lenses are overlaid on the same map. This leads to the identification of which of the known occurrences are most likely to be genetically related to banded iron formation.

Buffer zones are used in a wide range of GIS applications but often they are employed to model zones of influence from points, lines or areas. The first step in creating buffer zones is to calculate distances from the sources. The result of this distance calculation is that each grid cell will have a value assigned according to the distance from the source. In this case zones of influence are needed to model the occurrence of gold mineralization. The host rock of the gold mineralization is banded iron formation. It is assumed that this

lithology and the its vicinity represent areas with high potential for gold mineralizations. The width of the buffer zone will be arbitrarily set to 50 meters, i.e., a buffer zone of 50 meters around the banded iron formation lenses is generated. The result of the distance calculation is visualized by displaying the 50 m distance map, i.e., a binary raster map which contains only two pixel values: 1 for the areas within the buffer zone, and 0 for the areas outside the buffer zone.

**Area numbering and cross operation**

The second criteria to define areas with a high probability for gold mineralization, according to the hypotheses mentioned earlier, is the occurrence of shear zones/faults that intersect the buffer zones including the banded iron formations. To encode each shear zone/fault the automatic operation of area numbering is performed, i.e. a unique pixel value is assigned to each lineament. Then, the result of that operation is crossed with the buffer image of the BIF map. Finally, a map showing all the occurring combinations between the two maps is generated and in the cross table the numbers of the faults that intersect the BIF areas are listed. It may be of importance to identify which lineaments intersect the banded iron formation for the greatest length.

**Using 2-D tables for overlaying Faults and BIF**

Once the areas with high expected probabilities for gold mineralization have been mapped using both criteria (i.e., 50 meters proximity to BIF and occurrences of fault/shear zones), it is convenient to combine them into one probability map and to assign qualitative weights to all the possible combinations of pixel values between the two maps. To do such an analysis a two dimensional table can be used. Along one axis of the table there are the pixel values occurring in one map and along the other axis there are the values in the other map. Probability values may be assigned to the table in various ways, according to whether the faults intersect the banded iron formation and/or according to the length of intersection.

Finally, the result of this analysis can be combined with the geophysics and structural readings or any other data that might help in formalizing a geological interpretation. Important questions to answer might be the following. Are there patterns in the aeromagnetic data that relate to the banded iron formations? If yes, do these patterns also correlate with the strike of the bedding measured nearby? Can we structurally account for the offset or asymmetry in geophysical anomalies (e.g., by assuming that the total magnetic field is nearly vertical)?. Which of the known gold occurrences fall inside the final probability map? According to which criteria, do they fall in: (1) within the buffer of BIF or (2) along fault zones intersecting the BIF? Can we determine relative age relationships between dikes or faults? Are their orientations correlated to the orientations of interpreted lineaments? Can we determine the sense of displacements (sinistral or dextral ) along lineaments and dikes? What about vegetation cover and topographic relief in the areas of high probability? How much area of the high probability areas is covered by water? What is the distance of the high probability zones to the nearest lake larger then 1 square kilometre? Can we isolate a 10 kilometre scale mapping area for future study?

The following environmental questions can also be asked the database.
What are the possible land use issues that a summer field crew may encounter while in the field next year? What are the potential land use conflicts in the area of interest with

reference to points of landscape scenery, spectacular Wilbeforce Falls, fish spawning areas, caribou migration and the muskox population? Using the maps from the Parks Canada publication, which is also available, it is possible to overlay the natural resource maps with the gold prospect map (all the maps have to be georeferenced). What could the accuracy be for such an assessment considering the different map scales that have to be used?

### 7.1.5. Concluding remarks

This case study is directed not only towards exploration geologists but also towards environmental geoscientists. Much of the concepts discussed in this Section are of a very broad relevance. In particular, the emphasis is put on practical applications that go all the way from the fieldwork practice and the capture and storage of field data in a portable computer, to the drafting of maps, to the processing of remotely sensed data, to the design of predictive models and to their implementation in a GIS. The modern field work cannot anymore be isolated from spatial data analysis tools. Eventually, when all the preprocessing has been standardized, integration becomes synonym of predictive modelling. The two case studies that follow, deal with special exploration problems stemming out of complete research projects.

## REFERENCES


Agterberg, F.P., 1989, Computer Programs for Mineral Exploration, Science, v. 245, pp. 76–81.

Bonham-Carter, G.F., 1991. Integration of Geoscientific Data using GIS. p. 171–184. In: Geographic Information Systems. Vol. 2: Applications, D.J. Maguire, M. F. Goodchild and D.W. Rhind (Eds).

Fabry, A.G., Kushigbor, C.A., Wang, L., Baker, A.B. and Roscoe, S.M., 1989, Integration of remotely sensed and ancillary geological data for assessing mineral resources in the Bathurst Inlet area, N.W.T., Canada. In, Proc. 21st Application of Computers and Operation research in the Mineral Industry. Las Vegas, Nevada, Feb. 27 - March 2, 1989. Soc. of Mining Engineering Inc., Littleton, Colorado, pp. 1109–1120.

Fabbri, A.G. and Kushigbor, C.A., 1989, Relationships between integrated multisensor imagery, geophysical data and geological map patterns for mineral exploration East of Bathurst Inlet, N.W.T., Canada. In, Proc. IGARSS'89 Symp., Vancouver, B.C., July 10–14, 1989, pp. 1414–1420.

Henderson, M.N., Henderson, J.R., Jefferson, C.W., Wright, T.O., Willie, R., and Schaan, S., 1991, Preliminary geological map of the Hood River Belt (NTS 76K/N). Geological Survey of Canada Open File 2413.

Jefferson, C.W. , Henderson, M.N., Henderson J.R., Wright, T.O., Wyllie R., and Schaan, S., 1991, Geology of the Archean Hood river Belt, Northeastern Slave Structural Province, District of Mackenzie (NTS 76 K, N), NWT. Geological Survey of Canada, Open File 2484, pp. 193–197.

Roscoe, S.M., 1984, Assessment of Mineral Resource Potential in the Bathurst Inlet area, NTS 76 J, K, N, O Including the Proposed Bathurst Inlet National Park. Geological Survey of Canada, Open File Report 788, 75 pp.

Schetselaar, E.M., Brodaric, B. and de Kemp, E.A., 1993, Data Integration and GIS Modelling for the Field Based Earth Scientist: Workshop Tutorial. Ottawa, Canadian GIS Conference, March 1993 (unpublished manuscript).

Zoltai, S.C., Karasiuk, D.V. and Scotter, G.W., 1980, A Natural Resource Survey of the Bathurst Inlet Area, Northwest Territories, Parks Canada, 147 pp.

## 7.2. REMOTE SENSING AND GIS IN MINERAL EXPLORATION IN CENTRAL WESTERN SPAIN

M. A.Goossens
GeoConnaissance, Netherlands

### 7.2.1. Introduction

Ground based metal exploration is labour intensive and thus very expensive. Remote sensing and data integration techniques seem to offer a less expensive and attractive strategy to replace some of this effort, notably for identification of target areas in the screening phase of a new project. To realize this promise, however, one requires considerable knowhow in order to relate remotely sensed data to ground data, in such a way that the extracted signal forms a reliable predictive tool. This, in the first place, is because knowledge of the relations between geological setting, mineralizing processes, related anomalies, and available remotely sensed data form an essential basis for fruitful interpretation of remotely sensed data. In the second place, production of a reliable prediction requires considerable expertise concerning the technical aspects of data processing, and the evaluation of risks of misinterpretation and error estimation during the digital interpretation these data.

The intention of this case history is therefore twofold: (1) to show how specific geological and geochemical features, diagnostic for the geological setting of the mineralization, and the genetic processes, can be recognized using combined Thematic Mapper and airborne magnetic and radiometric data; and (2) to demonstrate how a variety of digital data processing techniques can be used to evaluate the quality of classification, and to enhance the reliability of the final prediction concerning the presence of mineralization.

### 7.2.2. Geological background

In this study we use the data for an area located in the province of Salamanca, central western Spain, where mineralization is predominantly associated with granite intrusion. All mineralization is found in the vicinity of the intrusive contact of Hercynian granitic rocks, with the Precambrian to Silurian metasedimentary rocks of the Complejo Esquisto Grauvaquico. Mineralization is mostly confined to the contact aureole. Most abundant are small, uneconomic, vein deposits, with tungsten, lead, copper, zinc and occasionally uranium or gold. The only deposit of potential economic significance is the Los Santos tungsten skarn.

The Los Santos (1.5 million tonnes, 0.79% WO3) garnet-pyroxene-scheelite skarn is formed by replacement of Cambrian limestones. The location of the skarn coincides with the area where the contact aureole reaches its widest horizontal extent, as shown in Fig. 7.7. The adjacent granitic intrusion, part of the Hercynian Central System Batholith, is strongly zoned, ranging from biotite-bearing granodiorites in its core to muscovite and cordierite-bearing monzogranites and aplites at its margin near the Los Santos deposit (Goossens, 1992). A map with the batholith zoning is in Fig. 7.7. The main geochemical variations within the intrusion can be explained with *in situ* fractional crystallization of the batholith, with gradual segregation of intercumulus melt causing the regional zoning.

During differentiation, the melt became progressively more peraluminous, reducing and becoming enriched in volatiles. The position of the skarn, adjacent to the most

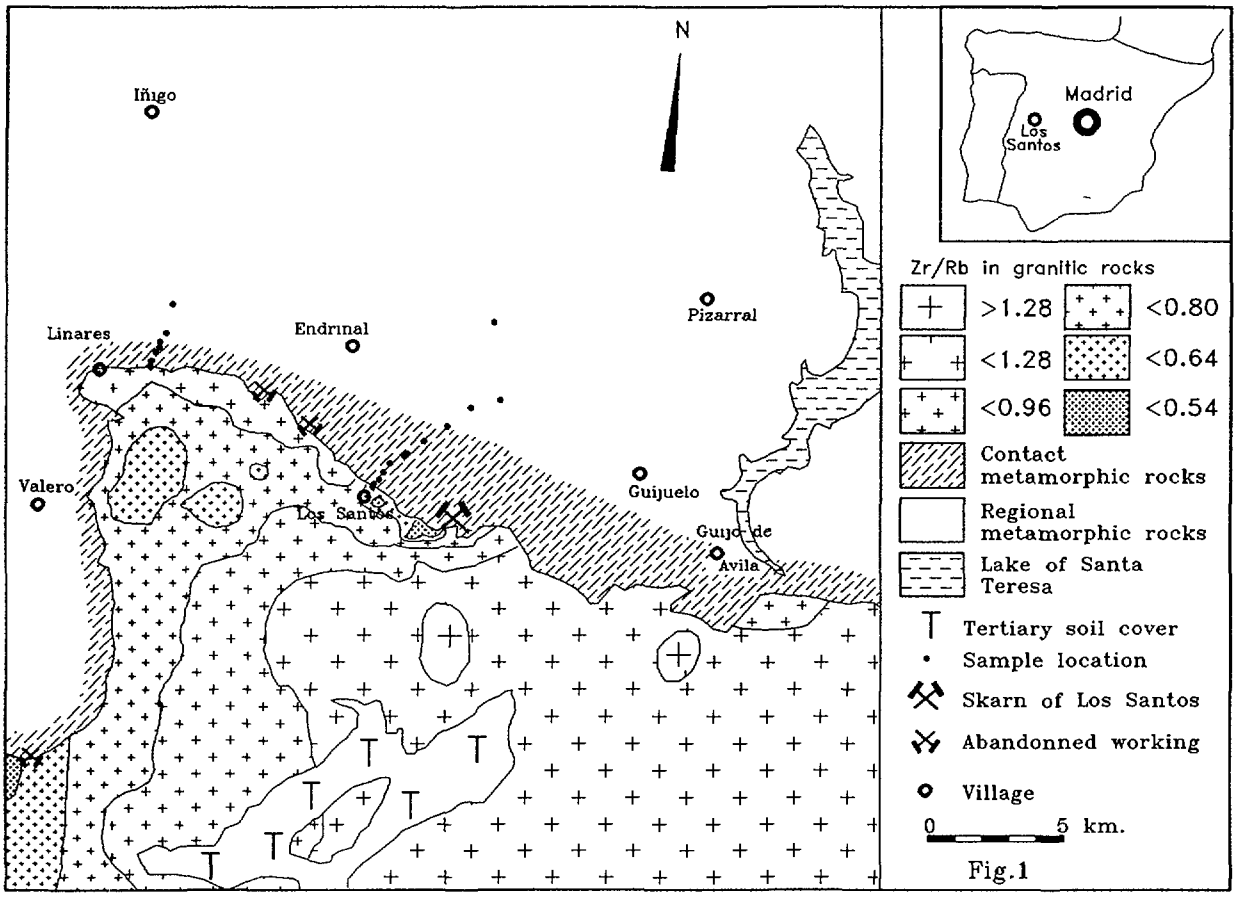



*FIG. 7.7. A geological and mineral deposit map of the study area.*

fractionated part of the intrusion, suggests that such mineralization is not only genetically related to the magmatic fractionation process, but additionally, it is to be expected in vicinity of the more felsic end members of a fractionation sequence. Local contamination processes have interfered with fractionation, obscuring the diagnostic trace elements, in particular those for Ba and Sr. Enhanced Ba and Sr levels, and depletion of Rb in the monzogranites adjacent to the Los Santos deposit, in combination with systematic Rb and Cs enrichment in the adjacent contact metamorphic rocks, suggest exchange of magmatic fluid. This exchange of fluids probably played a role during the formation of the skarn deposit.

### 7.2.3. Recognition of diagnostic features in TM and airborne geophysical data sets

Mineralization is often accompanied by anomalies which, although sharing some general characteristics, are highly diverse, and typical only for individual mineral deposits, as illustrated by the large amount of work , published on mineral deposits. Therefore it is more efficient to search for features, diagnostic for the setting of the type of mineralization that one is looking for, in particular in a situation like the one in the Los Santos area, where clear surface anomalies, such as gossans, are either absent, or were developed at such a small scale that remote detection is out of the question.

The main diagnostic feature that all mineralization in this area has in common, is the contact metamorphic setting. Interpretation of remote sensing data is therefore specifically directed towards the recognition of surface and subsurface expressions of contact metamorphism.

**Spectral classification of soils overlying the contact metamorphic rocks, using TM-band ratios**

A significant correlation has been found between the degree of contact metamorphism and related features (mineralogical composition, volatile content, $Fe^{++}/Fe_{total}$ ratio) of the rocks, and the chemical and mineralogical composition of the overlying soils (Goossens and Kroonenberg in press). From the exocontact towards the granite intrusion the proportions of kaolinite, illite and free iron (mostly present as goethite) in soils increase, while the proportions of chlorite and smectite decrease. Laboratory reflectance spectra of soil samples demonstrate that with the increase of the proportion of kaolinite, illite and free iron and the decrease of chlorite, the ratios of band2/band 3, band2/band5 and band2/band7 decrease while ratios of band3/band4, band4/band7 and band5/band7 increase. Those trends are shown in Fig. 7.8.

In the TM image, pixels that correspond with kaolinite- and goethite-rich soils are identified by elimination of other pixels during a process of stepwise masking various band ratio images. The argumentation for the pixel elimination is a follows:

(1) It is well known that the presence of dry and green plant material strongly affect the spectral signal. It was, in addition, found that reflectance is also significantly affected if more than 1.5% organic carbon is present. The first step therefore was to

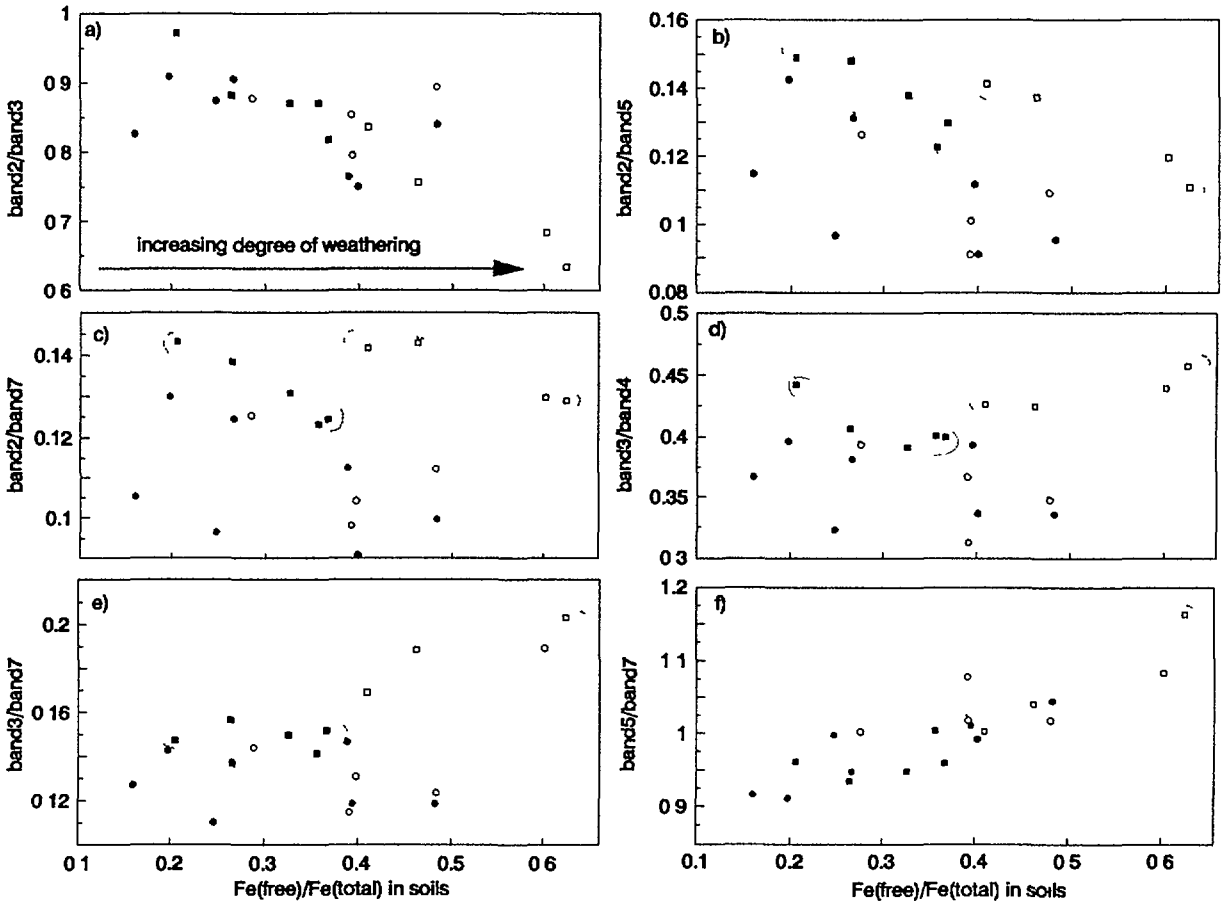



*FIG. 7.8. Variation of (laboratory derived) band ratios as a function of the free iron content (degree of weathering) in soils. Open symbols: kaolinite/chlorite >2. Closed symbols: kaolinite/chlorite <2. Circles: more than 1.5% of organic matter in soils.*

eliminate the effect of vegetation and organic material. The laboratory measurements yielded systematically higher ratios of [(TM5/TM 7)/(TM3/TM4)] for soils with more than 1.5% organic carbon, compared to those with less carbon. Therefore, resetting all pixels with high values of the ratio of [(TM5/TM7)/(TM3/TM4)] to zero, removes pixels that possibly correspond with soils rich in some sort of organic matter.

(2) It was found that an increasing amount of kaolinite, with simultaneously decreasing of chlorite amount in soils, was reflected in a decrease in the TM2/TM7 ratio. By resetting pixels (remaining after the previous step) with high values for this ratio, pixels that corresponded with chlorite-rich soils were removed from the image, leaving us with the most kaolinite-rich locations.

(3) Higher goethite contents in soils are reflected by higher values for the ratio of TM3/TM4. By resetting all (remaining) pixels with relatively low values for this ratio to zero, we were left with the pixels that have spectral properties most typical for soil material covering the contact aureole.

(4) Finally, a maximum likelihood classification was performed on original band ratio images, using the result of the masking as training set, in order to optimize the classification results.

**Detection of diagnostic textural differences: Land use**

Fieldwork has revealed that contact metamorphic rocks (hornfelses, spotted slates) were much more sensitive to weathering than regional metamorphic rocks (chlorite schists, garnet bearing biotite schists) of the Complejo Esquisto Grauvaquico or granites. As a result, soils overlying contact metamorphic rocks are, in this region, generally better and deeper developed, than those overlying other rocks. These properties are in turn reflected in a more diverse and intense land use within the contact aureole, compared to the remaining areas, resulting in a strong contrast in textural variability in the TM images. It was found that a very adequate way to map this contrast in textural variability of land use, was to pass a diversity filter over a single-band image. A diversity filter replaces the central pixel in a kernel by the total amount of different digital numbers in the kernel. Depending on the size of the kernel the diversity can range between 1 and 255.

**Classification of radiometric and magnetic data**

Systematic whole rock sampling has revealed diagnostic differences in radioelement (potassium, thorium and uranium) content for the different granite types and sedimentary rocks. Within the granite intrusion, radioelement content decreases with increasing degree of fractionation of the zoned intrusion. Compared to the granitic rocks, the metasedimentary rocks have low radioelement contents. Therefore, supervised classification of airborne radiometric thorium, uranium, and potassium images permitted adequate mapping of the different granite types and of the sedimentary rocks, which in turn enabled mapping of potentially contact metamorphic rocks.

The magnetic expression of the contact metamorphic rocks is different compared to that of the regional metamorphic sediments. Systematic whole rock analysis of metasediment samples, collected on traverses perpendicular to the intrusive contact (see Fig. 7.7), has demonstrated an increasing ratio of $Fe^{++}/Fe_{total}$ with increasing degree of contact

metamorphism. This changing ratio reflects reduction of $Fe^{+++}$ due during heating and recrystallization (Goossens, 1993), which in turn has led to a resetting of the remanent magnetism of the sedimentary rocks. This particular feature enables detection of regional metamorphic rocks that underwent contact metamorphic heating, using airborne magnetic data.

### 7.2.4. Analysis and integration of remotely sensed data

The clear relation between geology and the signal derived from remote sensing sources, as described above, suggests that it is rather simple to predict the position of possibly mineralized zones using only remote sensing data. However, this is not quite so, as illustrated by the common experience that the results of remote sensing analysis and data integration in mineral exploration may often be disappointing. A number of important reasons for this fact are identified:

(1) The basic relations between geology and the original data are often obscured by complex data manipulation, which obstructs the understanding of the final output.

(2) The risk of removing potentially important information during the interpretation of the data is considerable, and difficult to control.

(3) The decision criteria with respect to the correctness of classification and interpretation are often arbitrary and hard to describe in a quantitative manner.

(4) Error and quality assessment is often subjective, and therefore not reproducible under different circumstances.

No tailor-made solution can be presented for these problems, as this varies with each particular situation, and depends on the nature of a specific data set. Nevertheless a number of rules of thumb can be given, that enable careful and systematic processing of the data, so that a major part of these problems can be avoided.

This topic is a very important part of the study case, as a good understanding and feeling for the way the interpreted data should be handled, forms the backbone of a reliable result. However, it is not possible to discuss this matter adequately within a few lines of text. We therefore will briefly discuss this subject here, and deal with it very extensively in the practical part, allowing to experiment with the data, in order to become familiar with these very important concepts of data processing.

Concerning the first problem, it is important that the link between data, the interpretation and the geology is kept as direct as possible, and soundly based on the experiences obtained in the field. The data manipulation should be very straightforward and aimed at the recognition of specific features that are known to be diagnostic for the setting in which the mineralization is likely to occur. For example, the identified relation between metamorphic grade, soil composition and spectral signature is a feature that can be used for the detection of soils overlying contact metamorphic rocks, with a minimum of statistical or other kind of complicated data manipulation.

The second problem is perhaps the most difficult to deal with, in particular when a large, poorly documented area is investigated, as each step in the data processing, e.g., masking, filtering and classification, automatically implies loss of information. The

potential damage can be controlled by quantitatively describing each step, so that every manipulation is reproducible. If, in the course of the investigation it turns out that relevant data may have been lost, the steps can be repeated with modified parameters, and the results can still be compared.

With respect to the third problem it is important that the nature of the neighbourhood of a certain classified pixel is involved in the decision whether the classification is correct. For example, in the case of mapping 'kaolinite rich' pixels, typical for soils overlying contact rocks, using a TM image, one might want to use the concentration of classified pixels in a certain area as a criterion. For instance, we can apply the rule: "if a classified pixel is located close to many other similarly classified pixels, it is more likely to be correctly classified than a classified pixel that is far away form similar ones." This is illustrated in Figs 7.9 (a), (b). The binary image in Fig. 7.9 (a) shows the locations of pixels classified as 'kaolinite' using the TM image. In order to decide which of these pixels are correctly classified, and which ones are not, we passed a 5 × 5 rank order over the binary image, using a threshold of 17. The way such a rank order filter works is illustrated in Fig. 7.11.

The kernel, shown in Fig. 7.11 (a), covers a 5 × 5 grey-scale pixel neighbourhood, with values between 1 and 25. When ranked from low to high, the 17th value in this range of numbers is seventeen, and therefore the central value (16) will be replaced by 17. In the case of a binary image however, like for the classified image shown in Fig. 7.9 (a), the threshold determines the size of the removed clusters. Using a 5 × 5 kernel and a

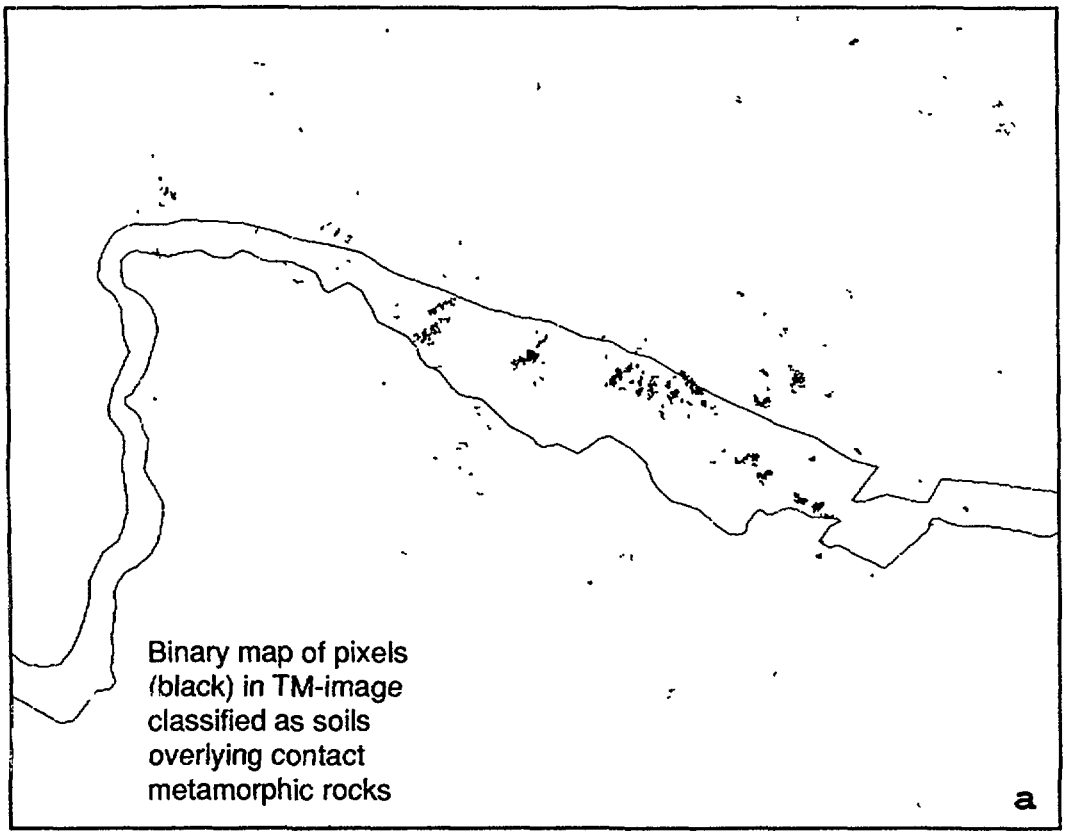
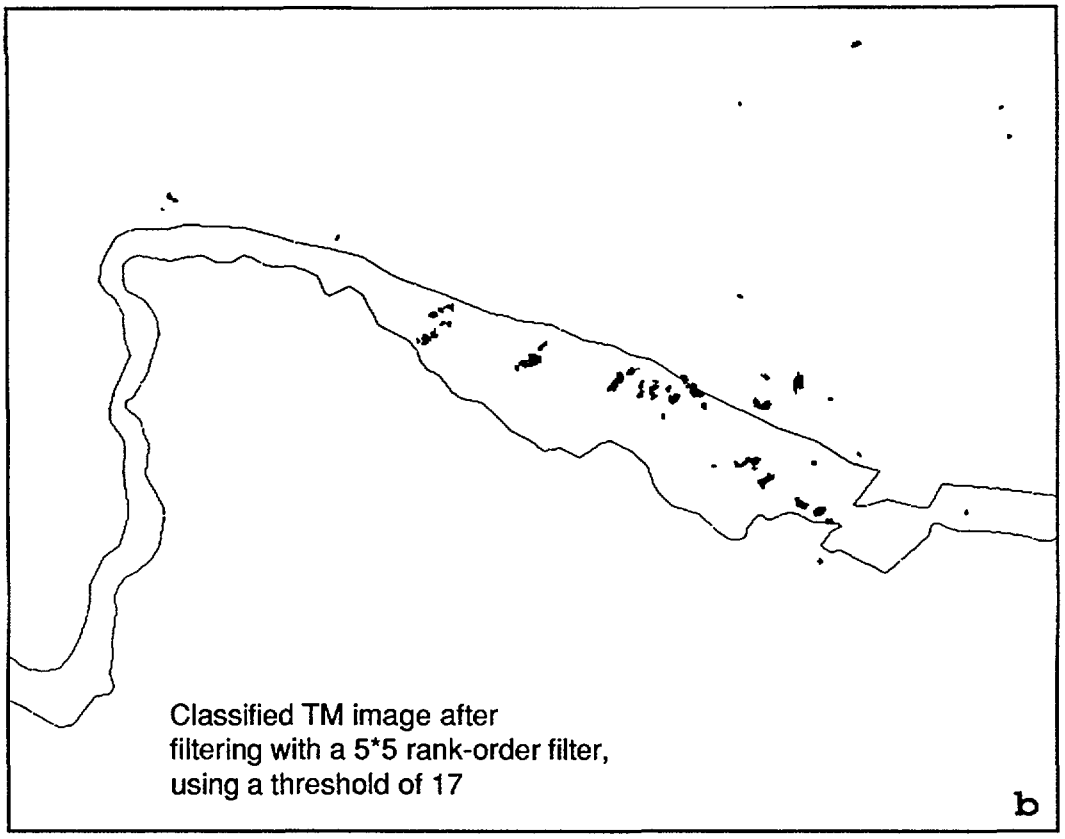



*FIG. 7.9. Binary maps of pixels classified as soils overlying metamorphic contact (a) and (b), and contour map of pixels classified as 'kaolinite' (c).*


Classified TM image after
filtering with a 5*5 rank-order filter,
using a threshold of 17
b

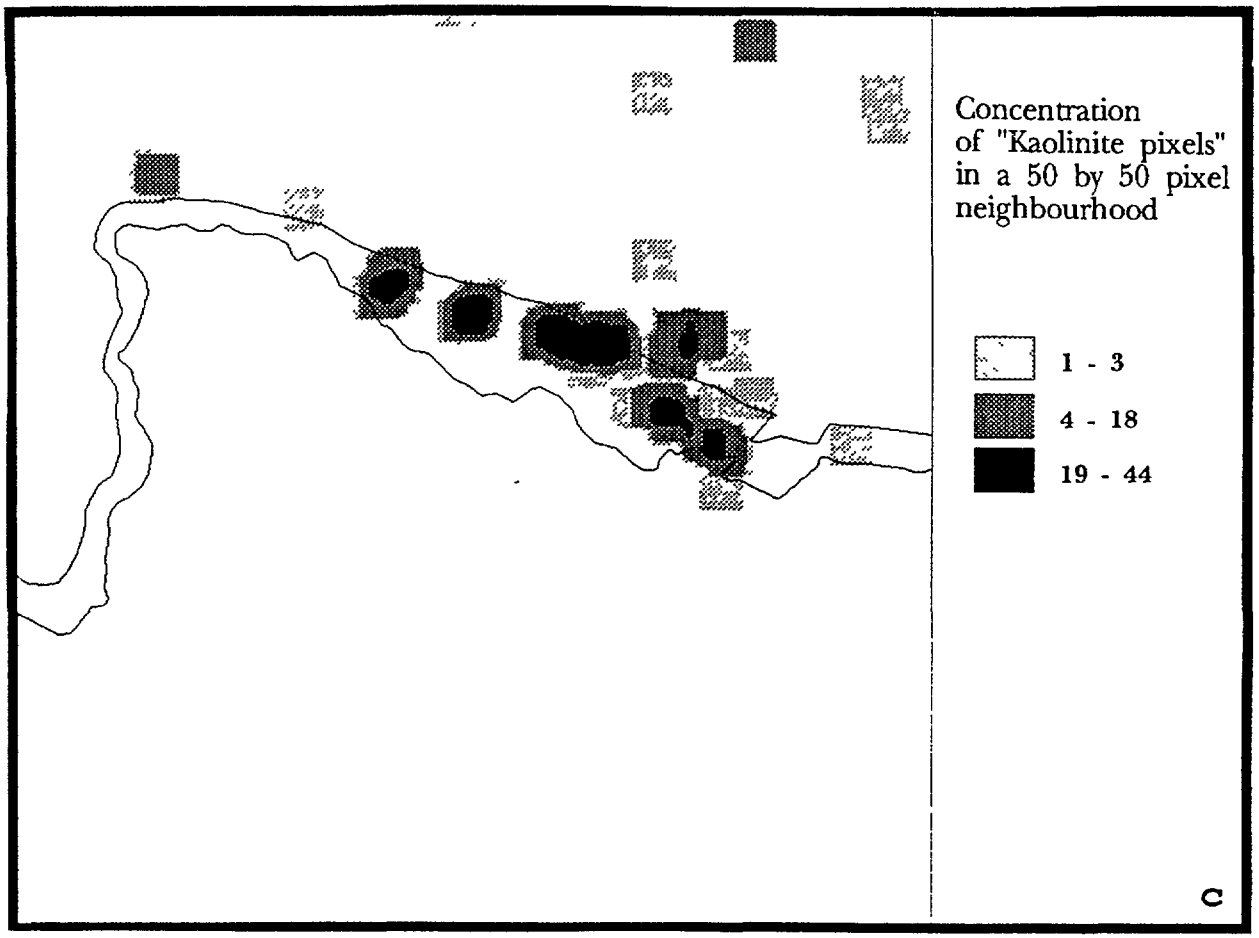

Concentration
of "Kaolinite pixels"
in a 50 by 50 pixel
neighbourhood
1 - 3
4 - 18
19 - 44
c

*FIG. 7.9. (cont.)*

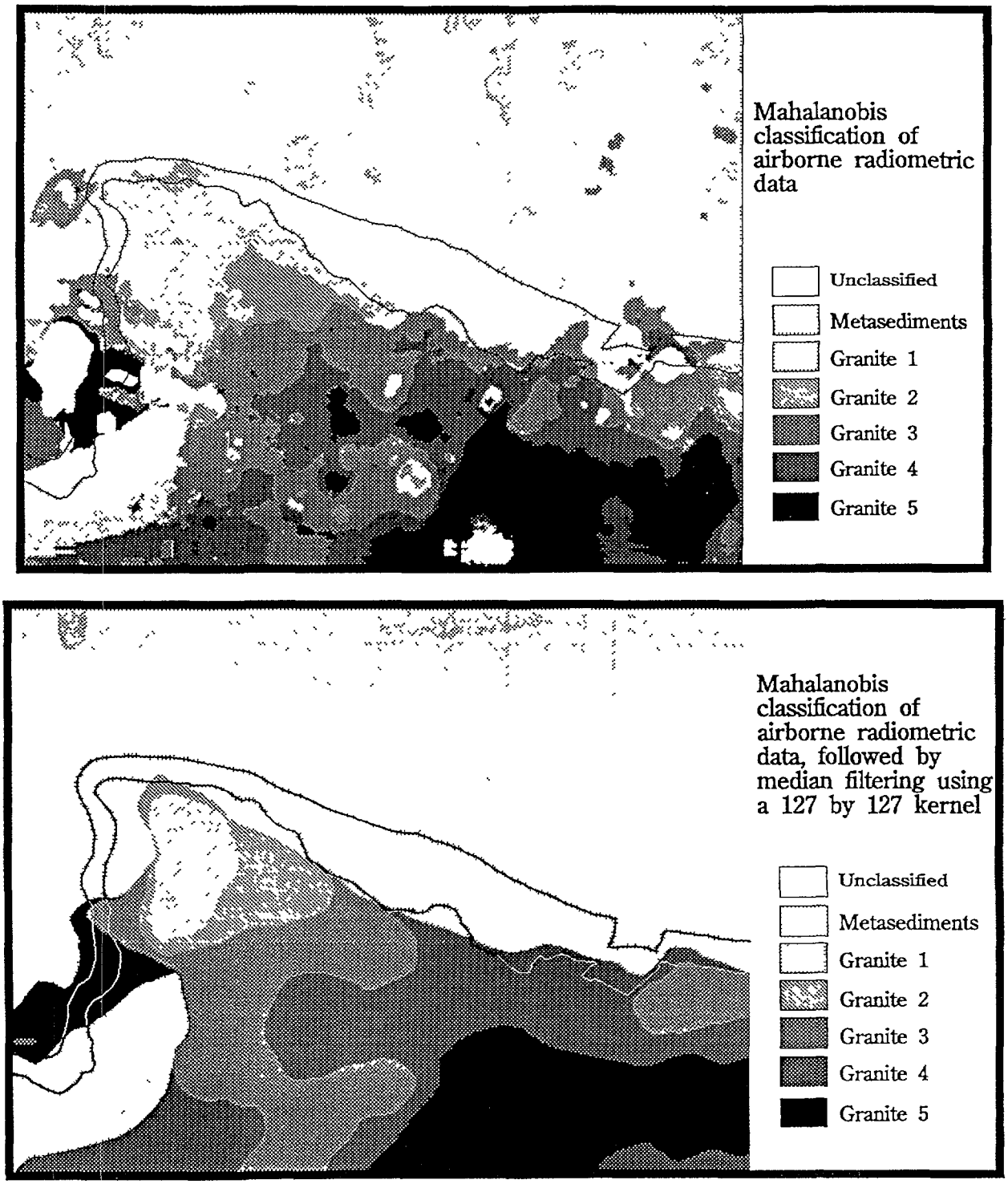



*FIG. 7.10. Rock classification on the basis of airborne uranium, thorium and potassium in the study area. (a) using a Mahalanobis classifier, and (b) the classified image after median filtering using a 127 × 127 kernel.*

threshold of 17 (see Fig. 7.11 (b)), a central 'binary 1' will be replaced by a central 'binary 0' if there are less than 8 'binaray 1s'. Consequently, classified pixels will be removed if a cluster consists of less than eight binary ones.

In the case of rock classification based on radiometric data, the size of a classified surface can be a valuable criterion. The larger the size of a homogeneously classified area, the

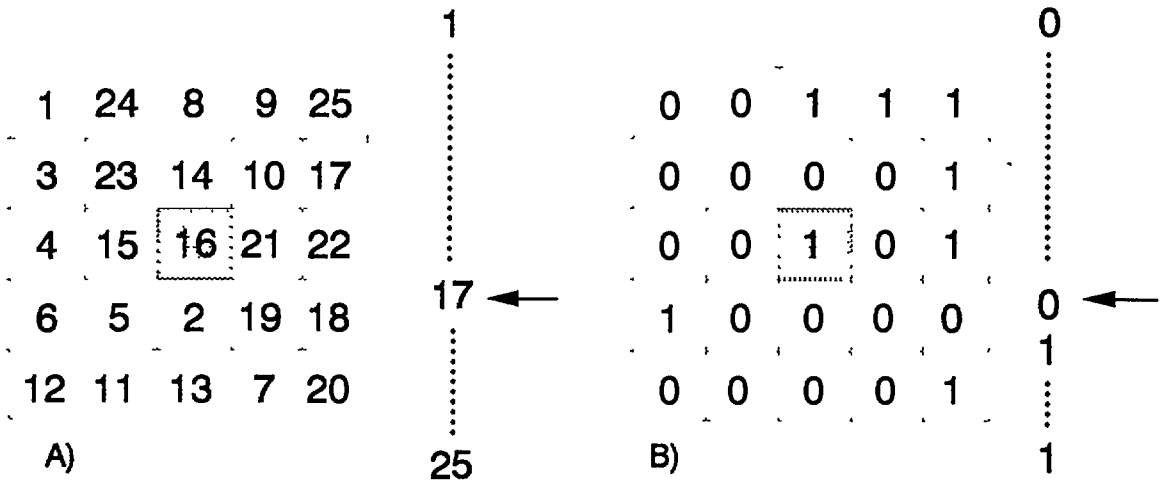



*FIG. 7.11. Demonstration of the effect of a 5 × 5 rank order filter, using a threshold of 17. (a) for a grey scale image; (b) for a binary image.*

greater is the chance that the classification is correct. This is exemplified in Fig. 7.10. In the classified radiometric image in Fig. 7.10 (a), it is likely that misclassification occurred, as small parts of the sedimentary terrain are classified as granite and vice versa. Possibly misclassified areas have been detected and corrected for by passing a 127 × 127 median filter over the image, as shown in Fig. 7.10 (b). A median filter replaces the central value in a kernel by the median of all values in the kernel. When the size of the filter is small, only small areas will be replaced by the value of the surrounding class. Increasing the size of the filter has the effect that larger areas will be regarded as misclassified. In the case of a 127 by 127 kernel, clusters smaller than 8065 pixels will be replaced.

In order to deal with the fourth problem, weights have been assigned to the degrees of uncertainty within an interpretation. For example, the pixels classified as 'kaolinitic' are not unique for the contact aureole (see Fig. 7.9 (b)), as they also occur elsewhere. However, their concentration is much larger inside of the contact aureole than outside. By contouring the classified image, a map is obtained that gives the concentration of pixels per surface unit. The range in concentrations was divided into three classes, as shown in Fig. 7.9 (c), and weights were assigned to each of the classes. As high concentrations are most typical for the contact aureole, the highest weights were given to these classes. Similarly, low weights are given to lower classes, as these are more typical for the area outside of the contact aureole. The actual weights that are assigned can be determined in various ways (Goossens 1992, Goossens, in press).

Subsequent integration of the interpreted and weighed data into a final map that expresses for each location the probability that mineralization will be present, is a very important step, because the result is that pixels, that were misclassified in one data set (for whatever reason), will not be confirmed by other data. They will therefore end with low probabilities in the final map. On the other hand, pixels that are most reliably classified will be highlighted, as they are confirmed by the multiple data sets.

## REFERENCES

Goossens, M.A., 1992, Petrogenesis of the mineralized granitic intrusion near Los Santos, Western Spain, and remote sensing and data integration as a tool in regional exploration for granite-related mineralization. PhD-Thesis; Geologica Ultraiectina 89, 164 pp.

Goossens, M.A., 1993, Integration of Landsat TM, Airborne magnetic and radiometric data, using a GIS, as an exploration tool for granite-related mineralization. Salamanca province, Western Spain. Nonrenewable Resources Vol. 2-1.

Goossens, M.A., Kroonenberg, S.B., in press, Spectral discrimination of contact metamorphic zones in Thematic Mapper images and its potential for mineral exploration, Province of Salamanca, Spain. Remote Sensing of Environment.

## 7.3. GOLD PREDICTION IN NORTHEAST FINLAND

V. Kuosmanen
Geological Survey of Finland

### 7.3.1. Introduction

This case history describes a study of integration of Landsat TM imagery, NIR aerovideo and aerogeophysical data for selection of gold prospecting targets in the Kuusamo area in NE Finland, as shown in Fig. 7.12. Part of this study was published in the IGARSS '91 meeting (Kuosmanen et al. 1991).

Gold occurs in a wide variety of geological formations, from sedimentary to volcanogenic. This requires that the prediction and exploration methods be quite diverse. This study outlines one possible way of integrating Landsat imagery with NIR aerovideo and with low-altitude geophysical measurements for pinpointing gold exploration targets.

In general, the methods used for Au exploration are anomaly-oriented. Geological, geochemical, geophysical or remote sensing anomalies which may indicate favourability for gold are mainly studied using single anomalies (Zeegers and Leduc 1991, Paterson and Hallof 1991, Coker and Shilts 1991). In the current study, interactive digital image processing was used for the integration of the geoscience data.

### 7.3.2. A review of the geology of the Kuusamo greenstone belt

The early Proterozoic Kuusamo volcano-sedimentary belt is part of a greenstone belt association extending from the Norwegian Sea to Lake Onega in Russia (see Fig. 7.12). Seventy percent of the belt is composed of sedimentary rocks with the remainder being mostly mafic volcanites, all unconformably overlying the Archean basement. Layered gabbro complexes and possible related volcanics have been dated as the oldest post-

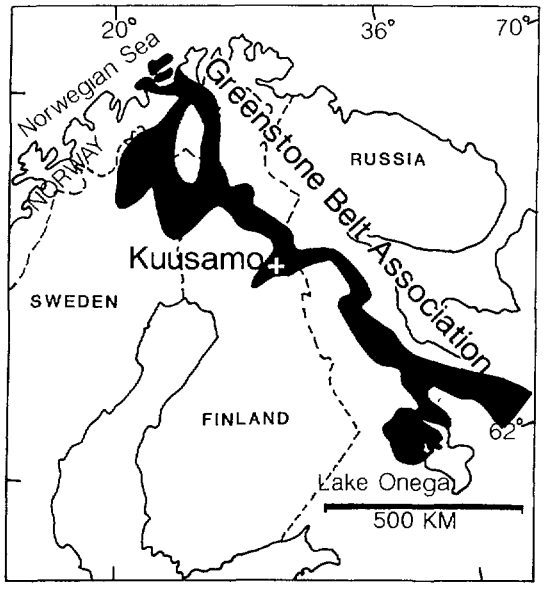



*FIG. 7.12. Location of the Kuusamo area and greenstone belt association modified after Pankka and Vanhanen (1989).*

Archean formation of the region, i.e. 2.44 Ga (Alapieti 1981). The western part of the belt was disrupted by granitoid intrusion during the Svecofennian Orogeny at about 19 Ga (Pankka and Vanhanen 1989).

An overview of the lithostratigraphy (Silvennoinen 1972, Pankka and Vanhanen 1989) of the Kuusamo area is shown in Fig. 7.13. The lowermost stratigraphic unit, Greenstone Formation I consists of mafic continental volcanic rocks. The Sericite Quartzite Formation overlies these volcanics. Most of the known Au bearing occurrences are in antiformal structures within the Sericite Quartzite. The Greenstone Formation II, a discontinuous, subaqueous basaltic flow, interrupts the sedimentary sequence. It is overlain by the Siltstone Formation. Concordant differentiated albite diabase sills and dykes are numerous and were intruded into undeformed, and in part, unconsolidated, sediments. Syenitic sills and dykes have also intruded the sedimentary units. The regional metamorphic grade in Kuusamo varies from lower to upper greenschist and amphibolite facies (Pankka and Vanhanen 1989).

### 7.3.3. Occurrence of gold

The occurrence of gold is related to the antiformal structure of the Sericite Quartzite Formation and to its hydrothermally altered parts, which are often form stockworks. About thirty such occurrences are known in the Kuusamo greenstone belt, of which

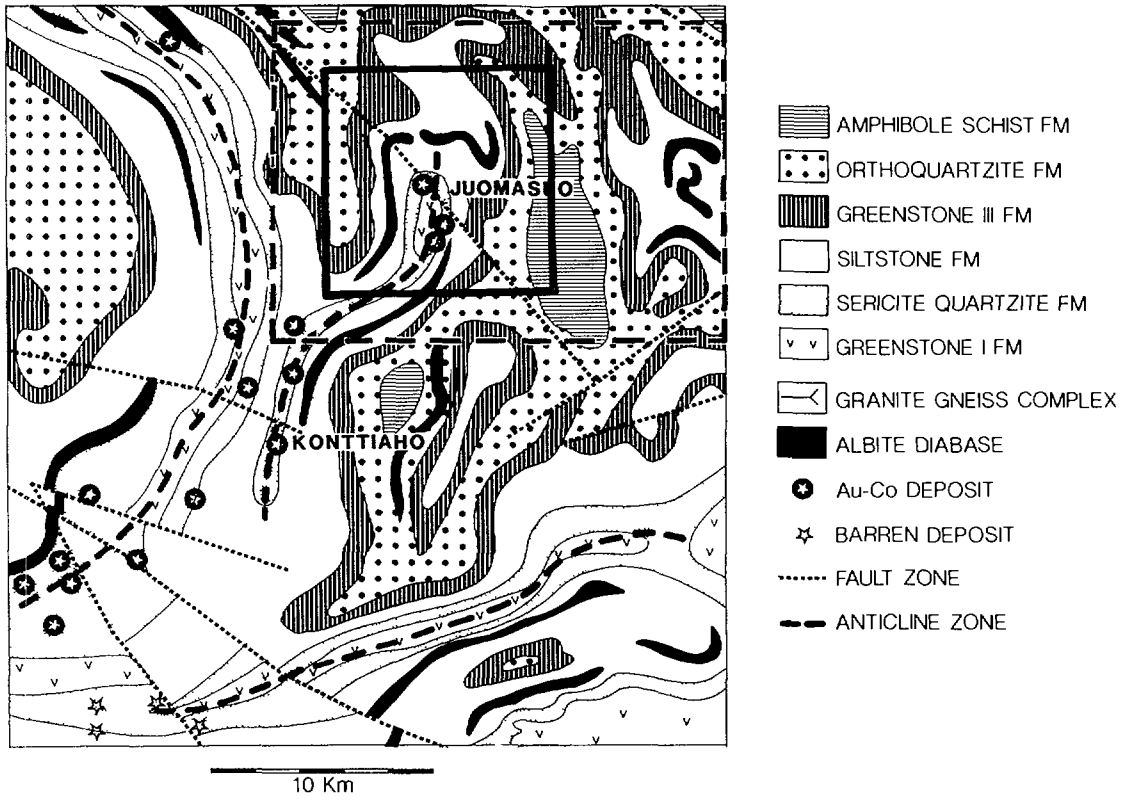



*FIG. 7.13. Lithostratigraphic map of the central part of the Kuusamo volcano-sedimentary belt and the main ore deposits. Modified after Pankka and Vanhanen (1989). Juomasuo and Konttiaho are the largest Au minerali-zations in the area. The 10 km × 10 km solid line square indicates an area of displayed data (see Fig. 7.20). The dashed line rectangle indicates the area of predicted targets (see Fig. 7.20).*

twenty are gold bearing. The average gold content in individual deposits varies from 0.5 to 10 ppm. The alteration zones are characterized by albite, carbonates, chlorite, quartz and sericite, and sulphide and U minerals. The gold mineralizations appear to be controlled by two fault systems. NE–SW trending axial rift faults control the antiform structures in the middle part of the belt (Fig. 7.13) and NW–SE faults control the location of mineral occurrences. The gold mineralization is epigenetic and early Proterozoic and appears to precede the peak metamorphism (Pankka and Vanhanen, 1989).

### 7.3.4. Data and strategy for the current study

This study is based mainly on airborne survey geophysical data, satellite imagery, NIR aeroimaging (Häme and Rantasuo, 1989) and field observations of gold occurrences. The strategy for the prediction is shown in Fig. 7.14.

The sites of 26 known Au-bearing mineralizations were NIR videoimaged (Häme and Rantasuo, 1989) from an altitude of 500–600 m. Flourishing vegetation possibly due to the fertilizing effect of carbonates could be perceived in proximity to some gold occurrences. These areas of flourishing vegetation were delineated and used as training data areas for the interpretation of Landsat infrared channels 4, 5 and 7.

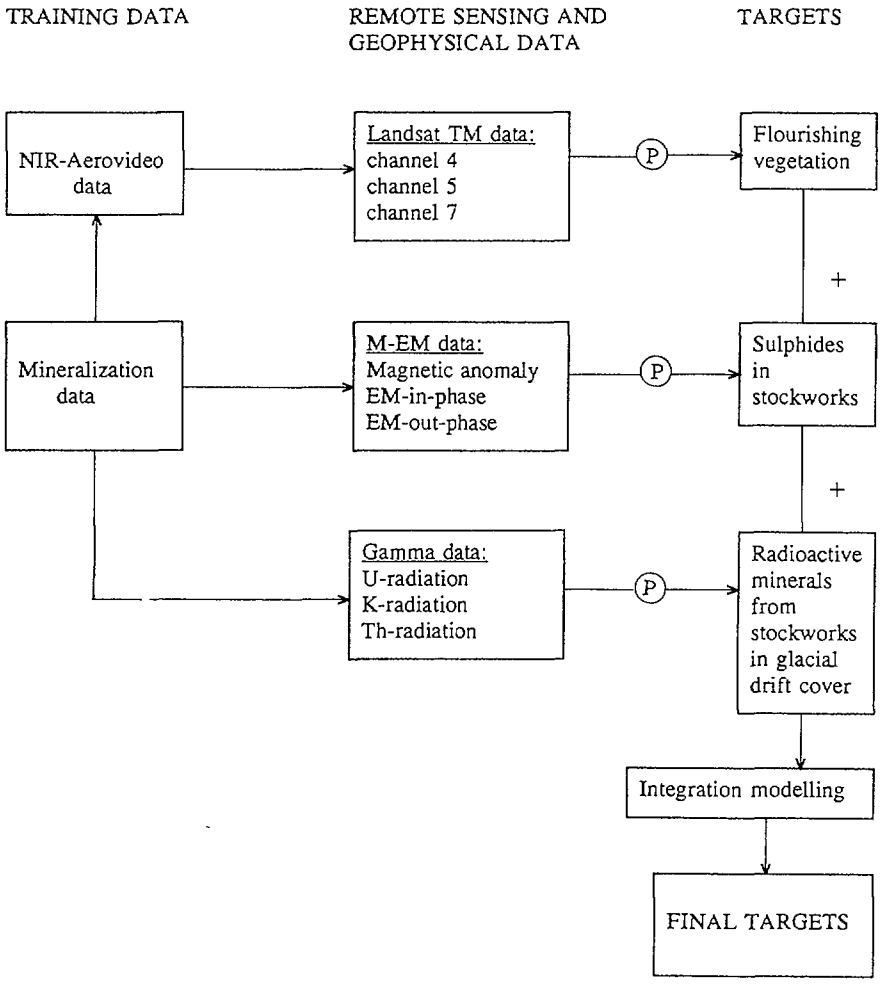



*FIG. 7.14. Strategy flow chart for the prediction for gold using several data sets. P refers to processing.*

The geophysical survey was made from an aircraft at an altitude of 30–50 m. Separation of the flight lines was 200 m and the sampling interval 12.5–50 m. Gamma-radiation U, Th, K [ppm], magnetic anomaly [nT], and Slingram electromagnetic in-phase [ppm] and out-of-phase [ppm] measurements were registered. Later in this text the magnetic and electromagnetic variables are called M-EM data and the radiometric K, U and Th variables are called gamma data.

### 7.3.5. Processing and integration

The Landsat TM images channels 4, 5 and 7 were geometrically corrected using bilinear interpolation to geographical map coordinates, as shown in Fig. 7.15. The aerogeophysical data were amplitude corrected along flight lines and linearly interpolated into regular 50 × 50 $m^2$ grid (see Fig. 7.15). These maps were then mosaicked and warped to form three 3-channel pixel images:

The indications of a gold bearing alteration pipe do not occupy the same geographical location: magnetic and electromagnetic anomalies mainly follow the distribution of sulphides in bedrock, as shown in Fig. 7.16, whereas gamma anomalies follow radioactive material glacially drifted from the pipe and mixed with glacial till, as shown in Fig. 7.17 (Plate 14). The vegetation anomalies are influenced by water-controlled dilution and precipitation patterns of carbonate, as seen in Fig. 7.16.

The corrected TM data (channels 4, 5 and 7) were used as the first group of features because of their correlation with the vegetation cover. M-EM data indicate the distribution of sulphides in the bedrock. Therefore these (corrected) M-EM variables were selected as the second group of features. The gamma data was used in the form of ratios U/K, U/Th and K/Th as the third group of features. These ratios were used because surface water effectively dampens gamma radiation but does not change the ratios. Processing for each group was done separately as it is indicated in the strategy flow chart shown in Fig. 7.14.

The 3-dimensional Landsat TM feature space is reviewed from direction (1, 1, 1) in Fig. 7.17 (Plate 14). The training data points in the centre of the picture represent Landsat TM pixels pinpointed by NIR-videoimaged areas of flourishing vegetation near the gold occurrences. Analogous sites, shown in Fig. 7.20 (Plate 16), for the vegetation were obtained by Nearest Neighbour classification of the Landsat TM data.

The 3-D feature space of the M-EM data was rotated into a direction where maximal separation of the training data points from all other M-EM observations could be seen, as shown in Fig. 7.18 (Plate 14).

The nearest neighbour areas from the most significant gold occurrences are indicated by the green spheres. The training data are indicated by coloured points, as shown in Fig. 7.18 (Plate 14). The directions of the rotated axes are as follows: MAGN = magnetic anomaly, RE = electromagnetic in-phase anomaly, IM = electromagnetic out-of-phase anomaly. Analogous sites (see Fig. 7.20, Plate 16) for gold occurrences after M-EM features were obtained also by Nearest Neighbour Classification. As was mentioned earlier, it was necessary to study ratios of gamma-variables because of the damping effect of water on the surface on the ground. The ratios U/Th, U/K and K/Th were first calculated and visualized. Therefore the classification of these features was made after ray-projecting all points of U, Th and K observations onto a plane perpendicular to the vector (1,1,1) i.e. plane U + Th + K = constant. The classification of the projected points (U/Th/K ratios) was carried out by Box Classification on that plane (Arkimaa

1992), as shown in Fig. 7.19 (Plate 15). The classification results of the gamma data are shown in Fig. 7.20 (Plate 16).

A simple integration of the classification results (targets) obtained (see Fig. 7.20) was carried out: Those areas where all analogies occur near ($\leq$500 m) to each other within sericite quartzite, were given the highest priority for field exploration.

### 7.3.6. Concluding remarks

Applicability of the predicted targets to mineral exploration seems to be relevant, especially, if the prediction models are delineated through visualizing the feature spaces

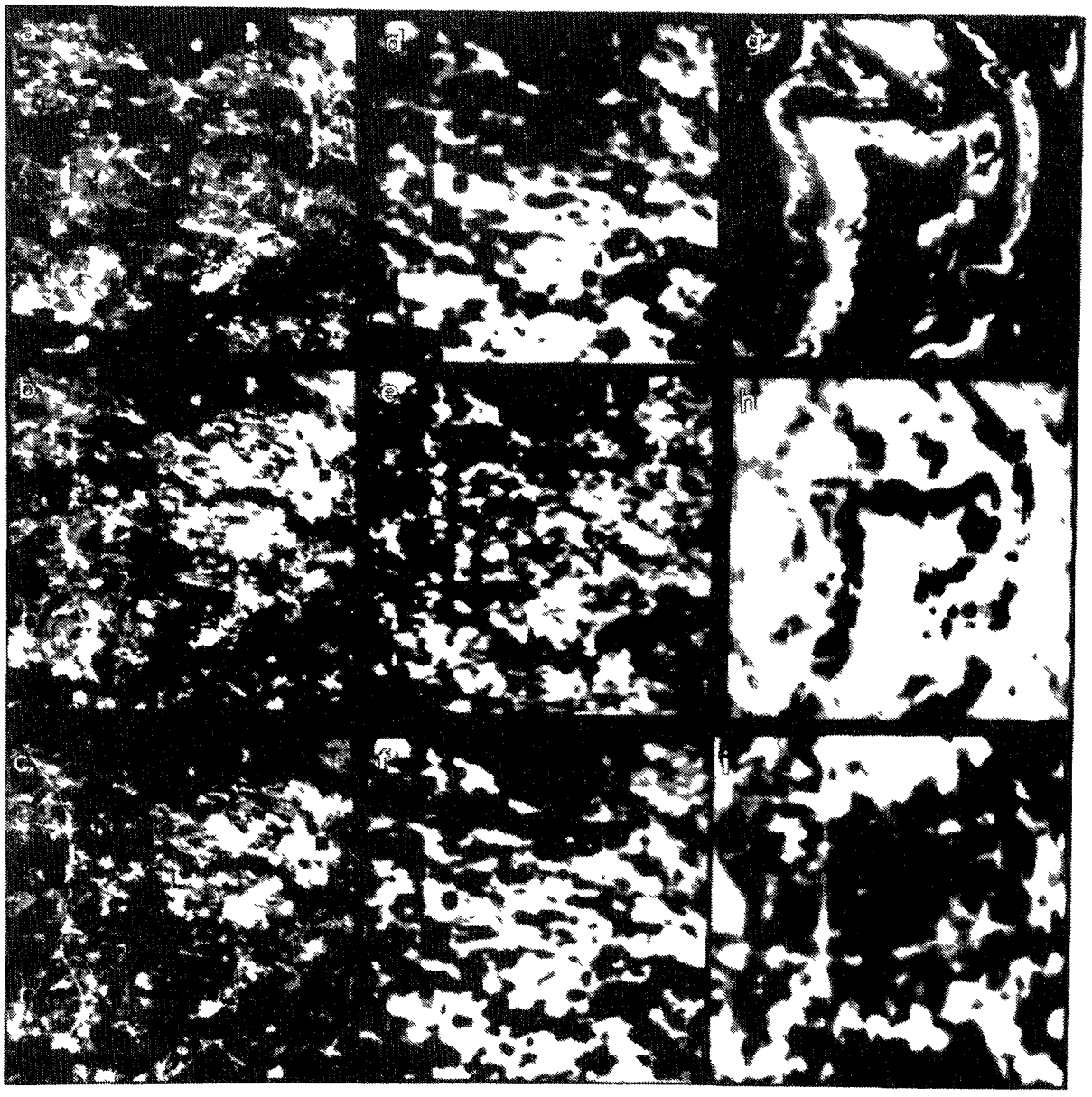


*FIG. 7.15. Examples of the preprocessed remote sensing and aero-geophysical data from the 10 × 10 km² area indicated in the lithostratigraphic map in Fig. 7.13. (a) Landsat TM channel 4, (b) TM channel 5, (c) TM channel 7, (d) potassium, (e) uranium, (f) thorium radiation, (g) magnetic anomalies, (h) electromagnetic in-phase anomalies, and (i) eectromagnetic out-of-phase anomalies.*

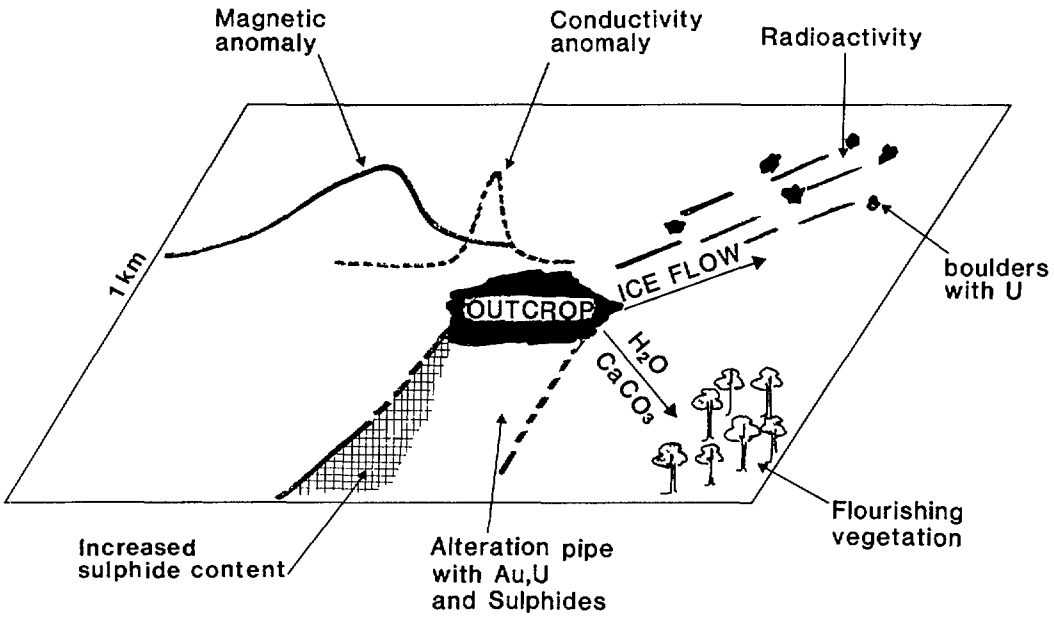



*FIG. 7.16. Various indications of a gold bearing alteration pipe stockwork. Notice that the different indications do not occupy the same geographical location.*

of the remote sensed variables. This ensures that the tailored model is concise enough to produce a minimum number of checkable pixels. However, the results are always affected by noise and erroneous signals.

Consequently, the resulting targets occur partly in rows and clusters. This corroborates the former idea of fracture control of gold occurrences (Kuosmanen et al. 1991). So far about 20 targets have been checked in the field. This resulted in finding two barren alteration deposits and two gold bearing mineralizations.

## REFERENCES

Alapieti, T., 1982, The Koillismaa layered igneous complex, Finland - its structure, mineralogy and geochemistry, with emphasis on the distribution of cromium, Geol. Surv. Finland, Bull. 319, 116 pp.

Arkimaa H., 1992, Study of airborne gamma ray spectrometric data in prediction of uraniferous mineralizations in the Kuusamo area, Northeastern Finland. In: New developments in uranium exploration, resources, production and demand, IAEA-TECDOC-650, pp. 29–33.

Cocker, W.B. and Shilts, W.W., 1991, Geochemical exploration for gold in glaciated terrain, in Foster (Ed.): Gold Metallogeny and Exploration, London, Blackie and Son Ltd, pp. 336–359.

Häme, T., Rantasuo, M., 1989, Shuttered camera-aerial colour video imaging in the visible and near infrared, Phot. Eng. and Rem. Sensing, Vol. 54, No. 12, pp. 1735–1738.

Kuosmanen, V., Arkimaa, H., Vanhanen, E., Talvitie, J., Laaksonen, J., 1991, Study of integrated Landsat imagery, NIR aerovideo and aerogeophysical data, for selection of gold prospecting targets in the Kuusamo area, NE Finland, IGARSS '91 Remote sensing: Global Monitoring for Earth Management, Espoo, Vol. 4, pp. 2069–2073.

Pankka, H. and Vanhanen, E., 1989, Aulacogen related epigenetic Au-Co-U deposits in northeastern Finland, in Autio, S. (Ed.): Geol. Surv. Finland, Special Paper 10, pp. 91–94.

Paterson, N.R. and Hallof, P.G., 1991, Geophysical exploration for gold, in Foster (Ed.): Gold Metallogeny and Exploration, London, Blackie and Son Ltd, pp. 360–398.

Silvennoinen, A., 1972, On the stratigraphic and structural geology of the Rukatunturi area, northeastern Finland, Geol. Surv. Finland, Bull. 257, 48 pp.

# 8. FUTURE TRENDS IN SPATIAL DATA PROCESSING

This handbook discusses various topics in the analysis of spatially distributed data giving special emphasis to mineral exploration. This chapter discusses likely future developments affecting geographical information systems. Table 8.1 lists the critical areas of development in spatial data processing.

TABLE 8.1. ACTIVE RESEARCH AREAS IN SPATIAL DATA PROCESSING

| | |
|---|---|
| 1. | Data acquisition, availability and transfer. |
| 2. | Data quality, representation and management. |
| 3. | Man-machine interaction, including visualization. |
| 4. | Data modelling, information integration, and decision support tools, including map generalization. |
| 5. | Spatial statistics and inference. |
| 6. | Training methods and tools, technology transfer. |
| 7. | New applications to resources, hazards and environments. |

1. The field of data acquisition will see extensive usage of geo-positioning systems or GPSs, to assign accurate spatial characterization to critical point observations. Low-cost scanners for document interpretation and storage are now available. New more specialized and powerfully designed sensors will be available for resource or disaster monitoring, while public and private archives will be providing basic cartographic data in standard formats on CD-ROMs and CD-Rs (Byte, 1993a) and through communication networks. Local area networks, LANs, and wide area networks, WANs, will enable sharing of databanks and of software.

2. Structured approaches to data collection will be developed where observations, measures of uncertainty, and interpretations will be clearly identified for multiple use of the data. Portable data management systems will be able to promptly accept the input of GPS and to provide descriptions of the positional accuracy of the inputs. New mass storage media on low-cost storage devices will be available to handle large hybrid databases with spatial data and non-spatial attributes, and integrated systems (Heley, 1991).

3. Interactive graphic user interfaces, GUIs, and discipline adaptable processing environments will make computers more widely used in the geosciences. Particular progress is now being made in three dimensional colour visualization and in 3-D GISs (Byte, 1993b; Buttenfield and Mackaness, 1991; Raper, 1989; Raper and Kelk, 1991; Pflug and Harbaugh, 1992; Turner, 1992; Sides, 1992). Visual computer languages, adaptable to different processing algebras, are now available for spatially distributed data (see the new Journal of Visual Languages and Computing, published since the year 1990; Treu, 1990; Egenhofer, 1990; Glinert et al., 1990; Myers, 1990; Batini et al., 1991).

4. Object-oriented programming (where objects are entities with attributes, relations and rules, including recursivity) and object-oriented databases, OOP and OODB, expert systems, decision support tools and the formalization of image processing strategies will enable to study more complex spatial relationships, e.g., in cartographic generalization (Buttenfield and McMaster, 1991; Muller, 1991), error propagation monitoring, image understanding, and information integration or data fusion (Burrough, 1991; Smith et al., 1987; Usery et al., 1988a, b, and Cress and Deister, 1990; Fedra et al., 1991; Albert, 1988; Smith and Yiang, 1991; Frank and Hegenhofer, 1992; Herring, 1992).

5. Spatial statistics and inference will lead to a better understanding of spatial processes and to the formulation of spatial predictive models (Ripley, 1991; Cressie, 1991; NRC, 1991).

6. Advanced training methods will be using on-screen hypermedia tools, mixing dialogue, iconics and image displays. Specialized applied training initiatives and institutes will facilitate worldwide technology transfer through international initiatives in research and development (Simard 1992a,b,; Geoscope, 1992; Fedra et al., 1991).

7. More GIS case studies will become available in the fields of renewable and non-renewable resources, natural and man-induced hazard, global and local environments (IDRISI, 1992; Fabbri et al., 1991; Schetselaar et al., 1990; van Westen, 1993; Schetselaar, et al, 1993).

Applications to mineral potential mapping (Singh et al., 1993), to data integration (Chung and Fabbri, 1993) and to environmental management (Fedra et al., 1991; Guttorp, 1991a, b) will lead to much more comprehensive approaches in the use of applied spatial statistics and decision systems.

Future challenges, however, will still remain in the area of image understanding and in our ability to express problems in computational form.

## REFERENCES


Albert, T.M., 1988, Knowledge-based geographic information systems (KBGIS): new analytic and data management tools. Mathematical Geology, v. 20, n. 8, pp. 1021–1035.

Batini, C., Cataraci, T., Costabile M.F., and Levialdi S., 1991, Visual Query Systems. University of Roma "La Sapienza", Department of Computer Science and Systems, Rap. 04.91, March 1991, 57 pp.

Beek, K.J. and Fabbri, A.G., 1991, GIS monitoring networks. Atti del Convegno sui Sistemi Informativi per la Difesa del Suolo e la Tutela del Territorio. Mondo Operaio, Roma, May 14-15, 1991, pp. 135–157.

Burrough, P.A., 1991, Intelligent geographical information systems – the case for formalizing what we already know. Proc. of Workshop on European Research in Geographical Information Systems (GIS), Davos, Switzerland, January 24–26, 1991, European Science Foundation, pp. 1–29.

Buttenfield, B.P. and Mackaness, W.A., 1991, Visualization. In Maguire, D.J., Goodchild, M.F. and Rhind, D., Eds, 1991, Geographical Information Systems: Principles and Applications. Harlow, Longman Scientific & Technical, pp. 427–443.

Buttenfield, B.P. and McMaster, R.B., Eds, 1991, Map Generalization: Making Rules for Knowledge Representation. Burnt Hill, Longman Scientific and Technical, 245 pp.

Byte, 1993a, Start the Presses (Create CD-ROMS on your Desktop), v. 18, n. 2, pp. 116–134, Feb. 1993.

Byte, 1993b, State of the Art: Visualization, v. 18, n. 4, pp. 120–148, April 1993.

Chung, C.F. and Fabbri, A.G., 1993, Representation of geoscience data for information integration. Jour. of Non-Renewable Resources, v. 2, n. 2.

Cress, J.J., 1990, Development and implementation of a knowledge-based geological engineering map production system. Photogrammetric Engineering and Remote Sensing, v. 56, n. 11, November 1990, pp. 1559–1535.

Cressie, N.A.C., 1991, Statistics for Spatial Data. New York, John Wiley and Sons, 900 pp.

Egenhofer, M.J., 1990, Interaction with Geographic Information Systems via spatial queries. Journal of Visual Languages and Computing, v. 1, n. 4, pp. 389–413.

Fabbri, A.G., Valenzuela, C.R., and Kushigbor, C. A., 1991, GIS training and research methods for resource management in developing countries. Proc. Canadian GIS Conference, Ottawa, Canada, March 1991, pp. 1024–1035.

Fedra, K., Winkelbauer, L., and Pantulu, V.R., 1991, Expert Systems for Environmental Screening: An Application in the Lower Mekong Basin. International Institute for Applied System Analysis, IIASA, Laxenburg, Austria, Research Paper RR-91-19, November 1991, 169 pp.

Frank, A.U. and Egenhofer, M.J., 1992, Computer cartography for GIS: an object-oriented view on the display transformation. Computers & Geosciences, v. 18, n. 8, pp. 975–987.

Journal of Visual Languages and Computing, 1990, Academic Press, London, published quarterly.

Glinert, E.P., Kopache, M.E., and McIntyre, D.W., 1990, Exploring the general-purpose visual alternative. Journal of Visual Languages and Computing, v. 1, n. 1, pp. 3–39.

Guttorp, P., 1991a, Spatial statistics in environmental science. In, National Research Council and National Academic Press, Spatial Statistics and Digital Image Analysis. Washington, D.C., pp. 71–86.

Guttorp, P., 1991b, Spatial statistics in ecology. In, National Research Council and National Academic Press, Spatial Statistics and Digital Image Analysis. Washington, D.C., pp. 129–146.

Healey, R.G., 1991, Database management systems. In, Maguire, D.J, Goodchild, M.F. and Rhind, D., Eds, 1991, Geographical Information Systems: Principles and Applications. Harlow, Longman Scientific & Technical, pp. 251–267.

Herring, J.R., 1992, TIGRIS: a data model for an object-oriented geographic information system. Computers & Geosciences, v. 18, n. 4, pp. 443–452.

IDRISI, 1992, IDRISI News: Version 4.0, TOPOS and UNITAR GIS Workbook Series Spring 1992, v. 4, n. 1, 6 p.

Muller, J.C., 1991, Generalization of Spatial Databases. In, Maguire, D.J, Goodchild, M.F. and Rhind, D., Eds, 1991, Geographical Information Systems: Principles and Applications. Harlow, Longman Scientific & Technical, pp. 457–475.

Myers, B.A., 1990, Taxonomies of visual programming and program visualization. Journal of Visual Languages and Computing, v. 1, n. 1, pp. 97–123.

NRC, National Research Council, 1991, Spatial Statistics and Digital Image Analysis. Washington, D.C., National Academic Press, 234 pp.

Pflug, R. and Harbaugh, J.W., Eds, 1992, Three Dimensional Computer Graphics in Modelling Geologic Structures and Simulating Geologic Processess. Proc. of Conference held in Freiburg, Germany, 7-11 October, 1990. Lecture Notes in Earth Sciences, v. 41, Berlin, Springer Verlag, 298 pp.

Raper, J.F., Ed., 1989, Three Dimensional Applications in Geographical Information Systems. London, Taylor and Francis, 189 pp.

Raper, J.F. and Kelk, B., 1991, Three-dimensional GIS. In, Maguire, D.J, Goodchild, M.F. and Rhind, D., Eds, 1991, Geographical Information Systems: Principles and Applications. Harlow, Longman Scientific & Technical, pp. 299–317.

Ripley, B.D., 1991, Statistical Inference for Spatial Processes. Cambridge, Cambridge University Press, 148 pp.

Schetselaar, E.M., Woldai, T., and Fabbri, A.G., 1990, GIS training methods: a geological case study. Atti del II° Workshop del GIAST - "Informatica e Scienze della Terra" - Sarnano, Italy, October 1990, pp. 202–211.

Schetselaar, E.M., Brodaric, B. and de Kemp, E.A., 1993, Data Integration for the Field Based Earth Scientist: Workshop Tutorial.

Sides, E.J., 1992, Modelling Three-Dimensional Geological Discontinuities for Mineral Evaluation. Unpublished Ph.D. thesis, Dept. of Geology, Royal School of Mines, Imperial College, London, 281 pp.

Simard, R. 1992a, A global change encyclopedia. Proc. of the Canadian Conference on GIS-92, Ottawa, Canada, March 24-26, 1992, pp. 949–960.

Simard, R. 1992b, Geoscope: an interactive global change encyclopedia. Proc. of Meeting of the International Space Year, March 1992, Munich, Germany, 5 pp.

Singh, P.K., Grunsky, E.C., Vander-Flier Keller, E. and Keller, C.P., 1993, Porphyry copper potential mapping using probabilistic methods and geographical information systems in British Columbia. Proc. of the GIS '93 Symposium, Vancouver, British Columbia, Canada, February 1993, pp. 381–394.

Smith, T.R., Pequet, D.J., Menon, S. and Agarval, P. 1987, KBGIS-II: A Knowledge-based Geographical Information System. Intl. Jour. of Geogr. Inf. Systems, v.1, n. 2, pp. 149–172.

Smith, T.R. and Yiang, J., 1991, Knowledge-based Approaches in GIS. In, Maguire, D.J., Goodchild, M.F. and Rhind, D., Eds, 1991, Geographical Information Systems: Principles and Applications. Harlow, Longman Scientific & Technical, pp. 413–425.

Treu, S., 'Conceptual distance' and interface-supported visualization of information objects and patterns. Journal of Visual Languages and Computing, v. 1, n. 4, pp. 369–388.

Turner, K.A. (Ed.), 1992, Three-Dimensional Modelling with Geoscientific Information Systems. Dordrecht, Kluver Academic Publisher, 443 pp.

Usery, E.L., Altheide, P., Deister, R.R.P. and Barr, D.J., 1988a, Knowledge-based GIS techniques applied to geological engineering. Photogrammetric Engineering and Remote Sensing, v. 54, n. 11, pp. 1623–1628.

Usery, E.L., Deister, R.R. and Barr, D.J., 1988b, A geological engineering application of a knowledge-based geographic information system. Proc. American Congress on Surveying and Mapping/American Society for Photogrammetry and Remote Sensing Annual Convention, The World in Space, Volume 2: Cartography, St. Louis, Missouri, 1988, pp. 176–185.

van Westen, C.J., 1993, Application of Geographic Information Systems to Landslide Hazard Zonation. Enschede, Netherlands, ITC Publication 15, 245 pp. (unpublished Ph.D. Thesis).

# PLATES

## PLATES

# PLATE 1

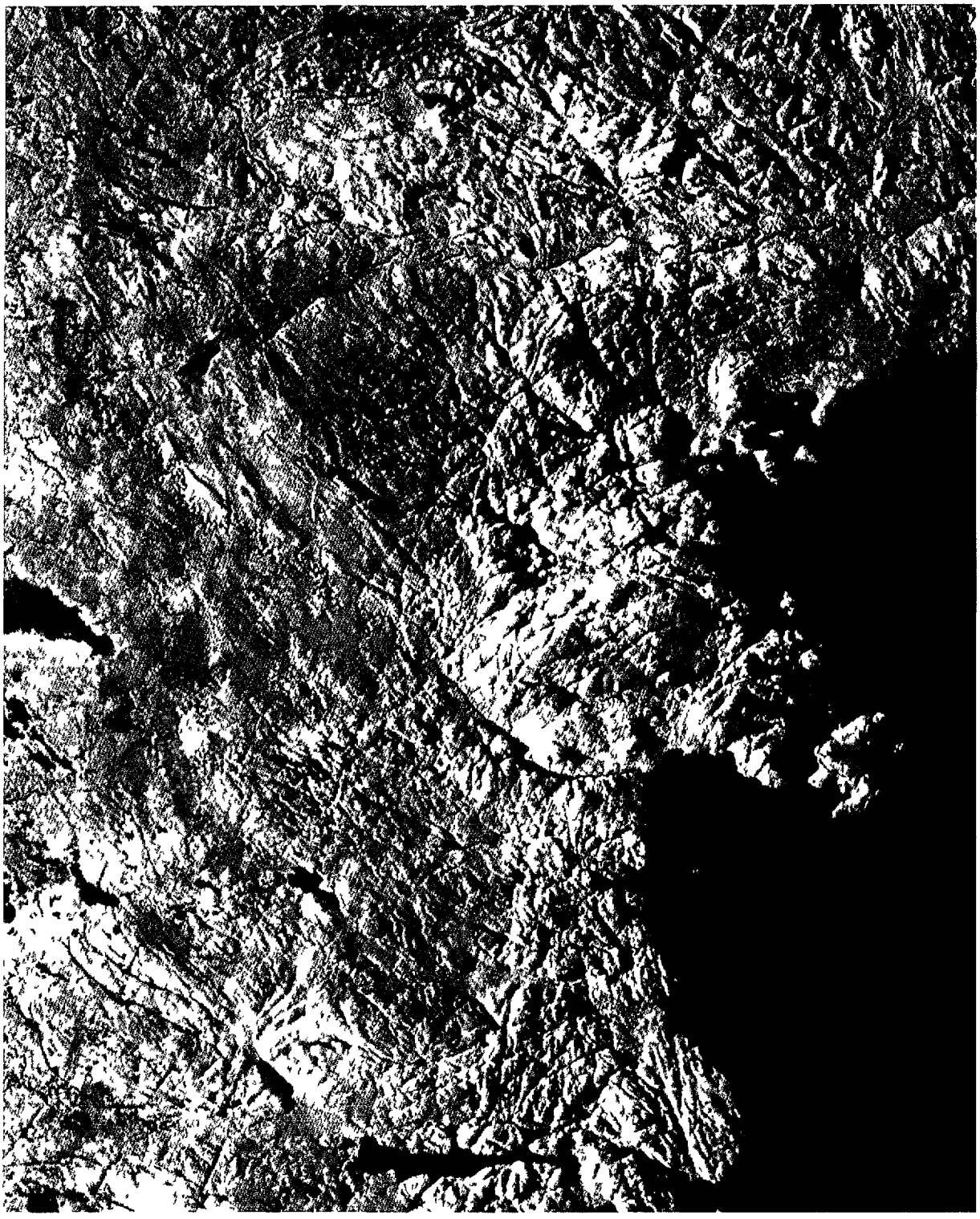


*FIG. 3 6. IHS display of an airborne radiometric image combined with a radar image for the Marathon area on the north shore of Lake Superior. The circular yellow feature is the Coldwell alkaline intrusive complex. (By courtesy of D. Graham, Canada Centre for Remote Sensing.)*

# PLATE 2

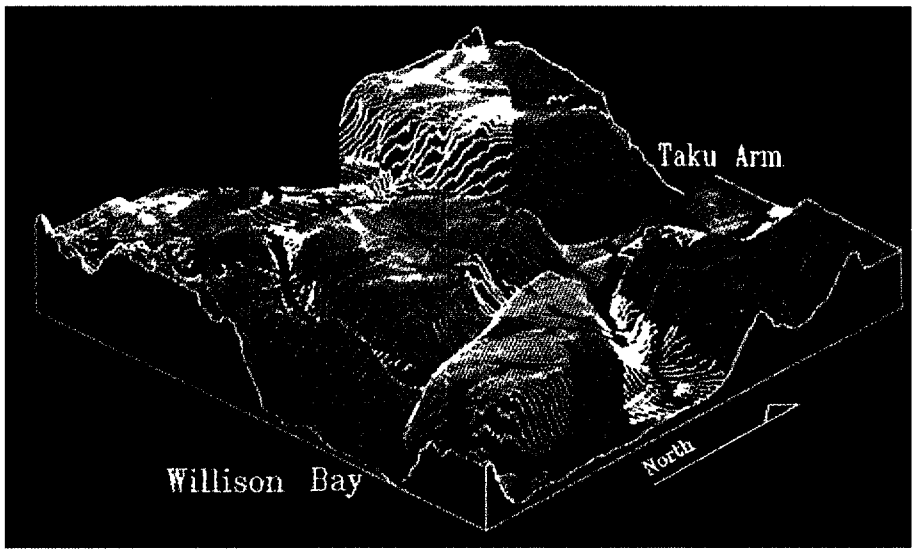



*FIG. 3.10. Geological map 'draped' over wire frame perspective image. This is an effective way to visualize the relationship between two different images. (By courtesy of A. Rencz, Geological Survey of Canada.)*

# PLATE 3

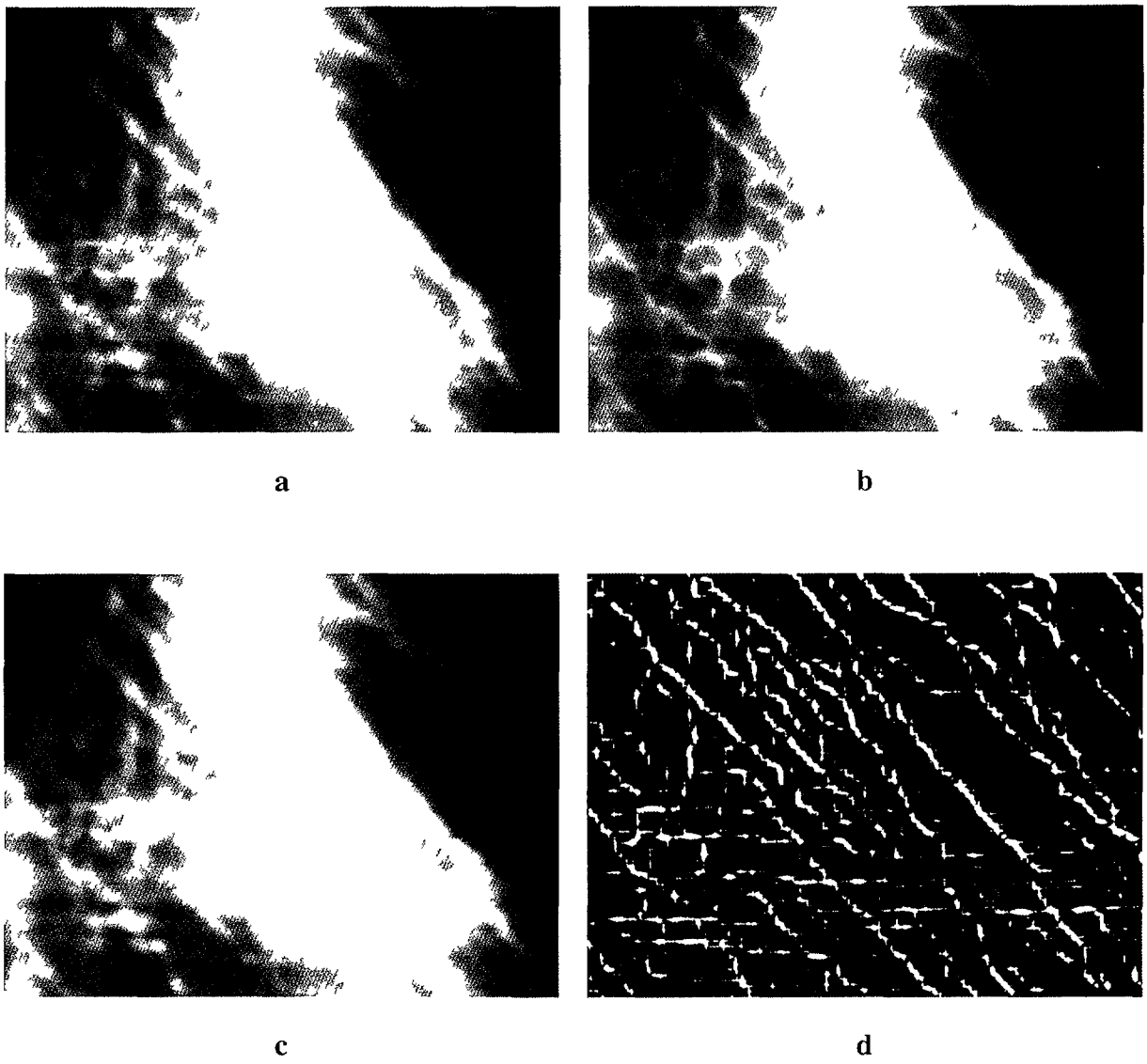
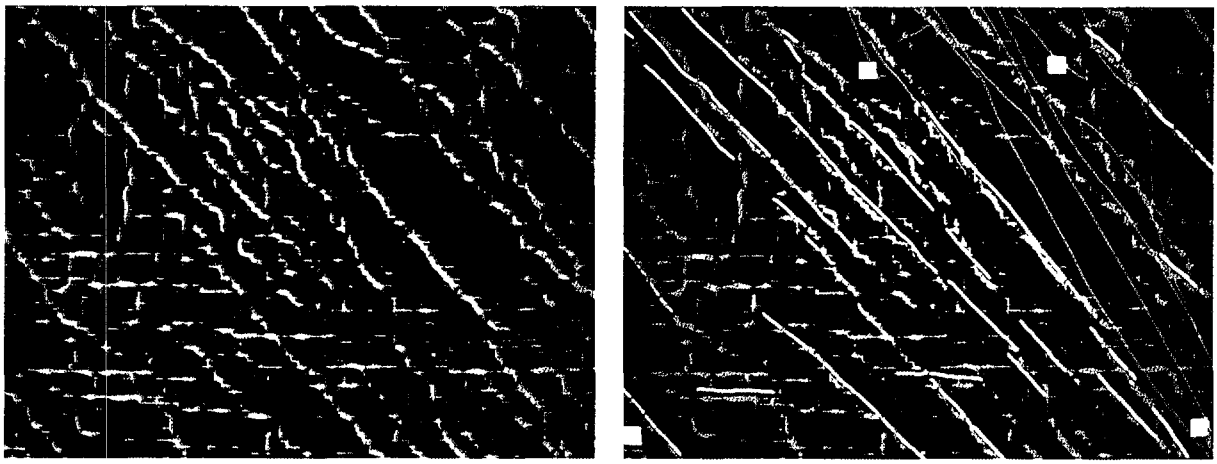



*FIG 4.8 Grey level morphological filtering of the image of aeromagnetic anomaly over Bathurst Inlet, Northwest Territories, Canada. (a) Subimage of 800 pixels by 600 lines,with resolution of 100 m per pixel (b) Erosion by a circular structuring element of diameter of 9 pixels (minimum within the kernel) (c) Opening of the image in (a) by dilatation of the image in (b) by the same structuring element (maximum following a minimum within the kernel) (d) The stretched image of the subtraction of the opened image in (c) from the original image in (a)*

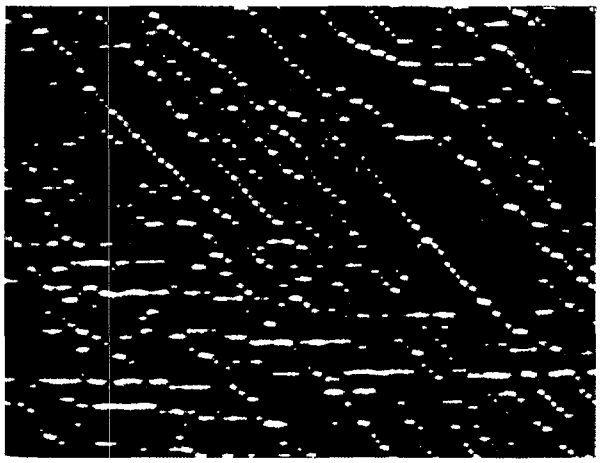
# PLATE 4



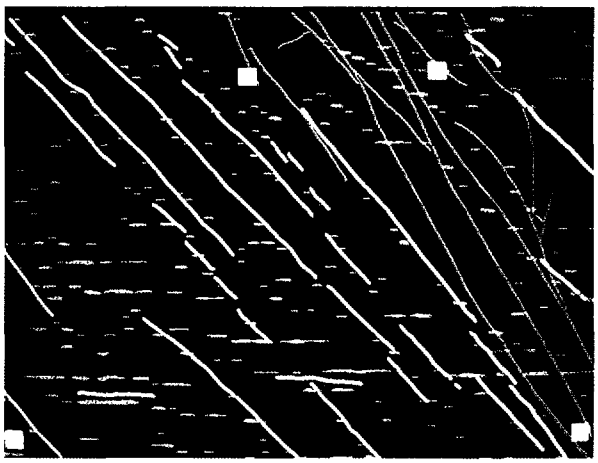
a



b



c



d

*FIG. 4.9. Further processing of the aeromagnetic anomaly image in Fig. 4.8. (a) The top-hat transformation of the image in Fig. 4.8 (a), computed by thresholding the opened image in Fig. 4.8 (d) for pixel values equal or greater than 2. (b) Overlay of the image of digitized dikes (white), faults (red) and mineral occurrences (coloured squares) from a published geological map. (c) The top-hat transformation of the image in Fig. 4.8 (a) using a vertical linear structuring element of length 7 pixels. (d) Overlay of the digitized structural and occurrence data with the image in (c) after opening it by a horizontal linear structuring element of length 9 pixels (colours are the same as in (b)).*

# PLATE 5

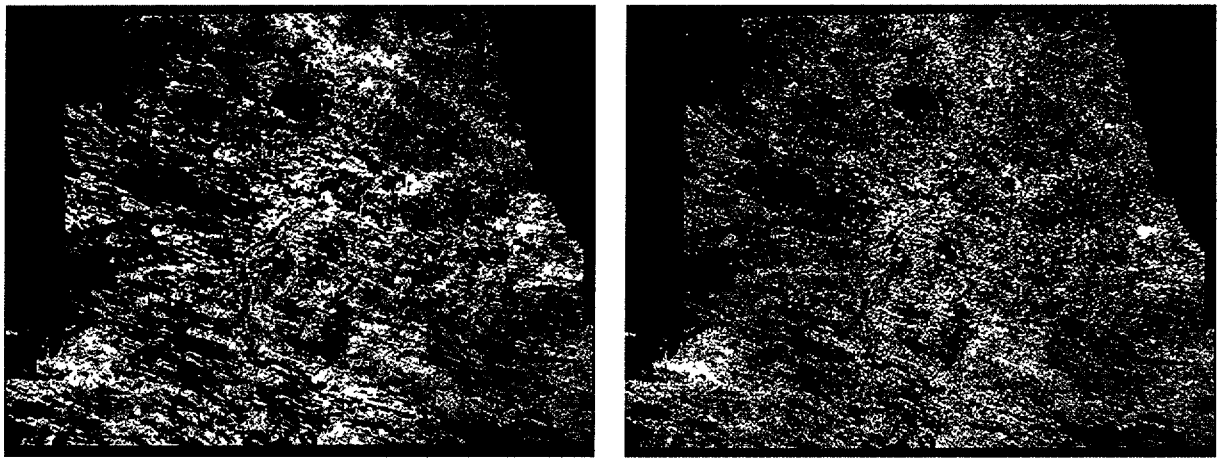


*FIG. 4.13. Additive colour composite of airborne survey gamma radiation data. (a) Original K, U and Th. (b) Principal components of PC 1, PC 2 and PC 3 in red, green and blue, respectively.*

# PLATE 6

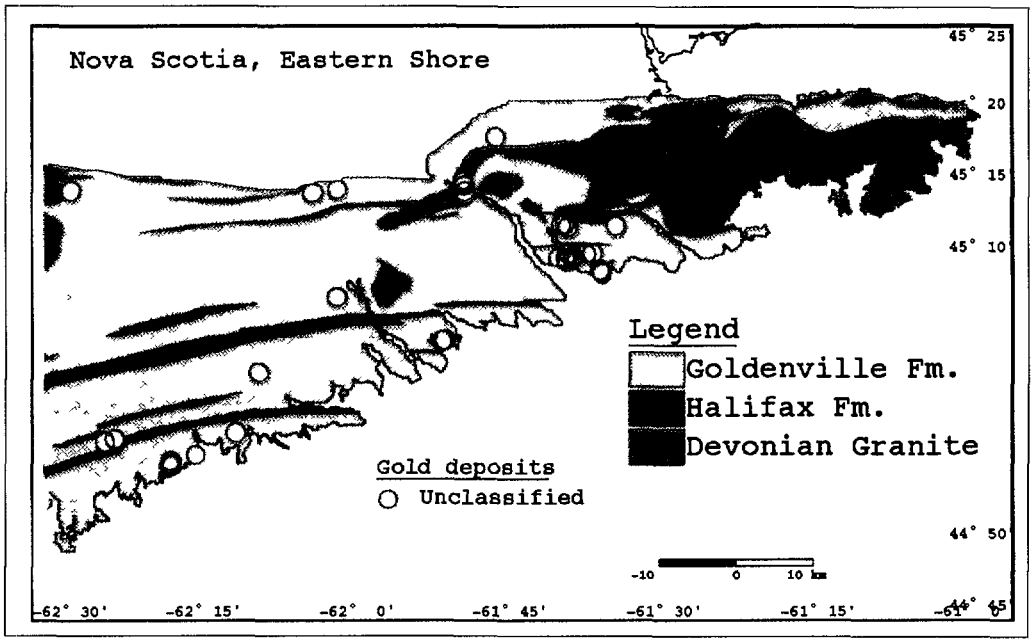



*FIG. 5.8. Geological map showing the locations of 35 gold deposits. The Goldenville and Halifax Formations are lower Paleozoic turbidites. The gold occurs in quartz veins, often associated with arsenopyrite and stibnite. In this area the deposits are all in the Goldenville, but further west some occurrences are in the Halifax Formation. The sediments are folded with axes approximately oriented ENE-WSW.*

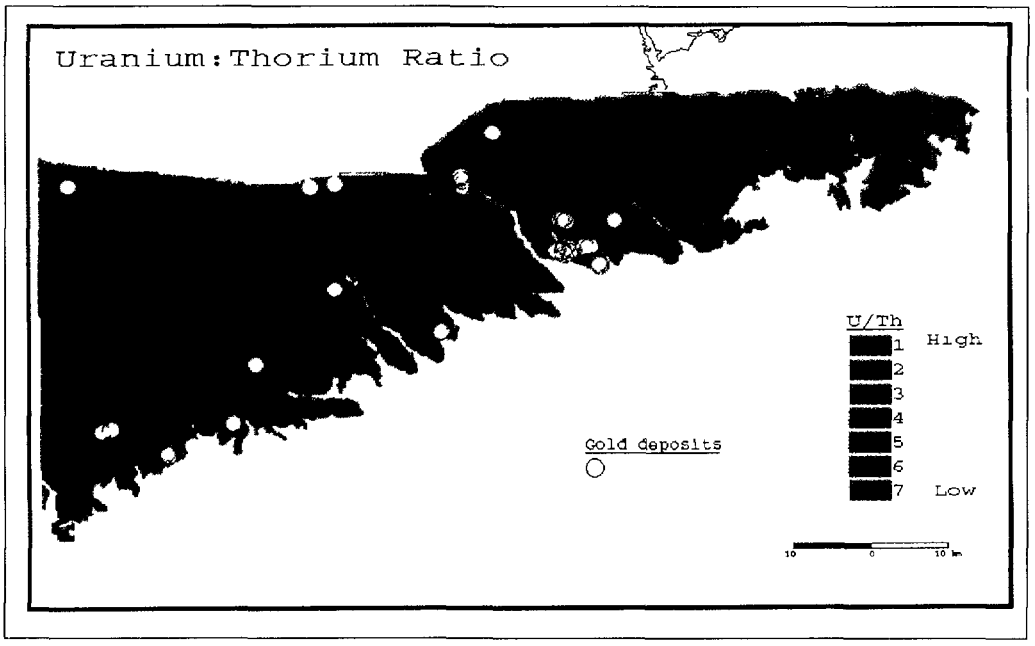



*FIG. 5.9. Airborne radiometric map, interpolated from flight line data. The Devonian granites have an elevated U–Th ratio. No gold deposits occur in the granite.*

# PLATE 7

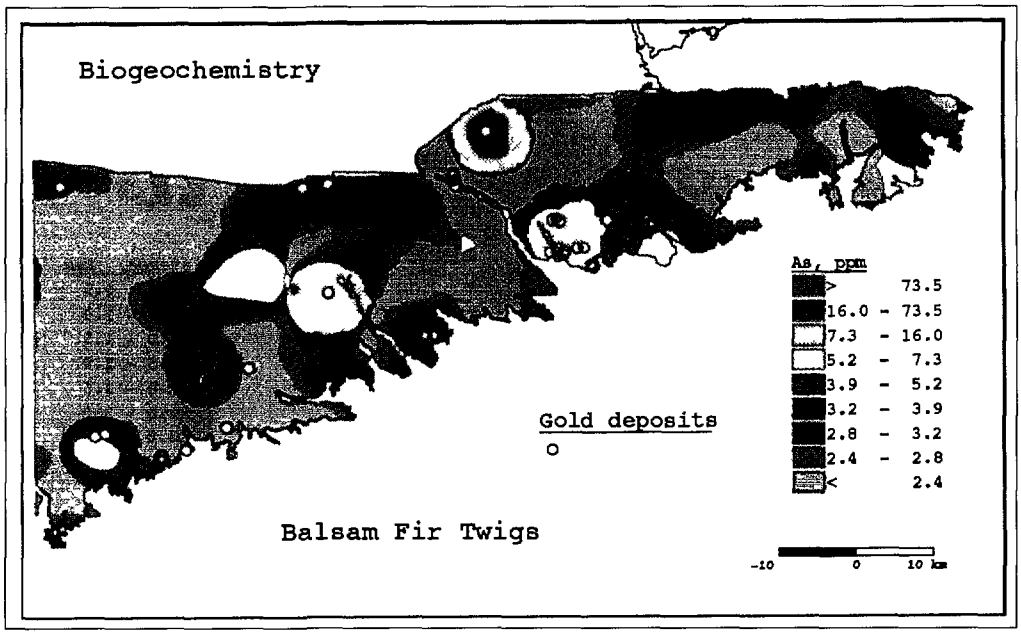



*FIG. 5.10. Biogeochemical map showing the mapped levels of As from about 500 samples of Balsam Fir twigs. The original point values have been interpolated on to a raster image and are represented on a percentile scale. Arsenic is an important pathfinder element for Au in this area, and regional biogeochemical maps have been used for Au exploration.*

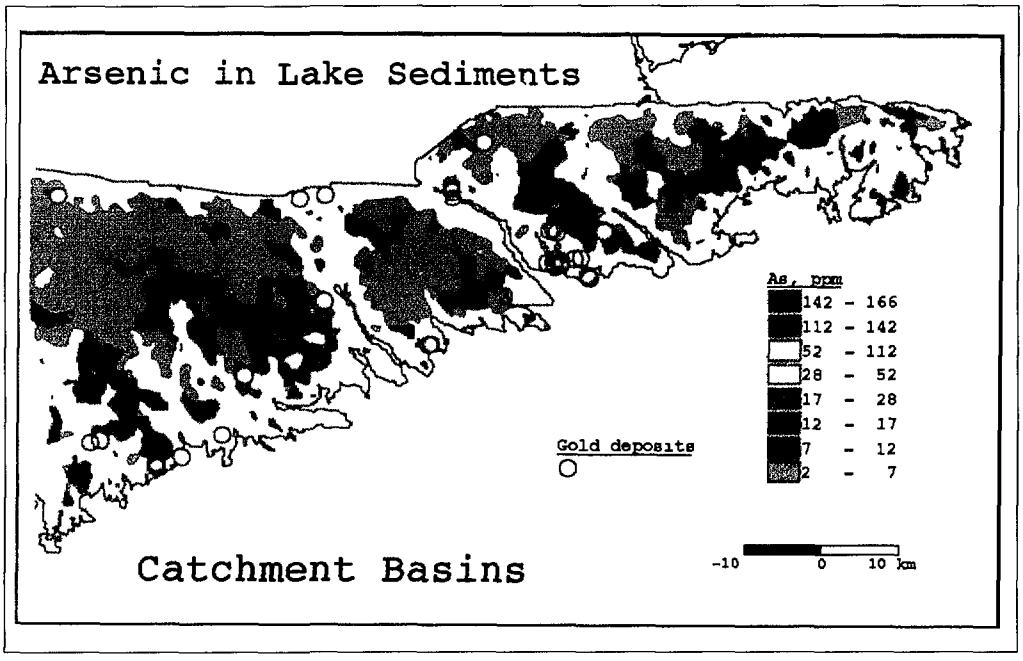



*FIG. 5.11. Lake sediment geochemical map. Each lake sediment sample is represented by a catchment area. The map was produced by assigning the As value for each sample to the corresponding catchment basin, then classifying the data into percentile levels.*

# PLATE 8

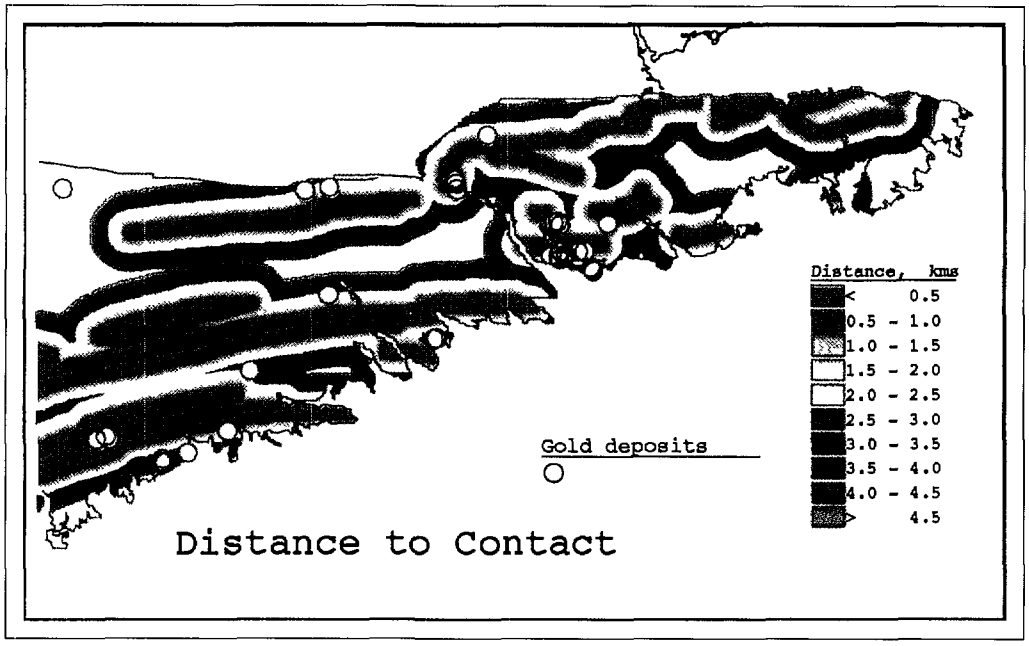



*FIG. 5.12. Proximity to contact between Goldenville and Halifax Formations. Some research has suggested that this contact is significant in the localization of gold deposits. The map was produced by extracting the specific contact from a vector file of the geological map, followed by successive dilation or buffering to produce a raster proximity map.*

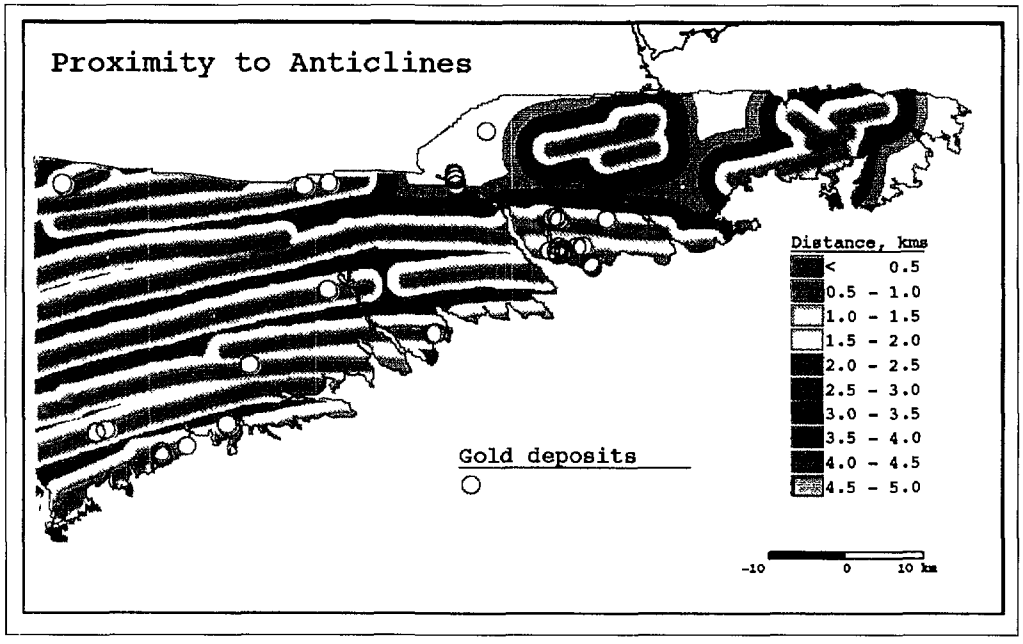



*FIG. 5.13. Proximity to the surface traces of anticlinal axes. Many of the gold deposits occur close to fold axes. The map was made by table digitizing the axes from a published map, then successively dilating the lines and producing a raster map of anticline proximity.*

# PLATE 9

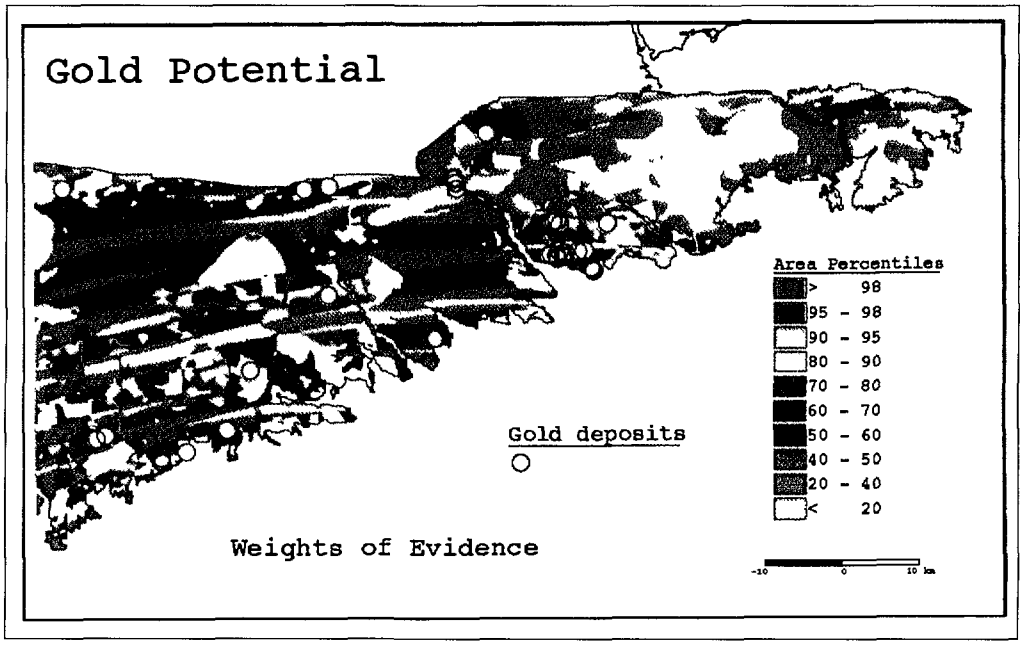



*FIG. 5.14. Gold potential derived by combining six binary input maps with weights of evidence. The reclassification of multi-class to 2-class maps, and the calculated weights are shown in Table 5.5. The classification of the potential map is made on the basis of area percentiles, making comparison with the results from logistic regression to be discussed later more straightforward.*

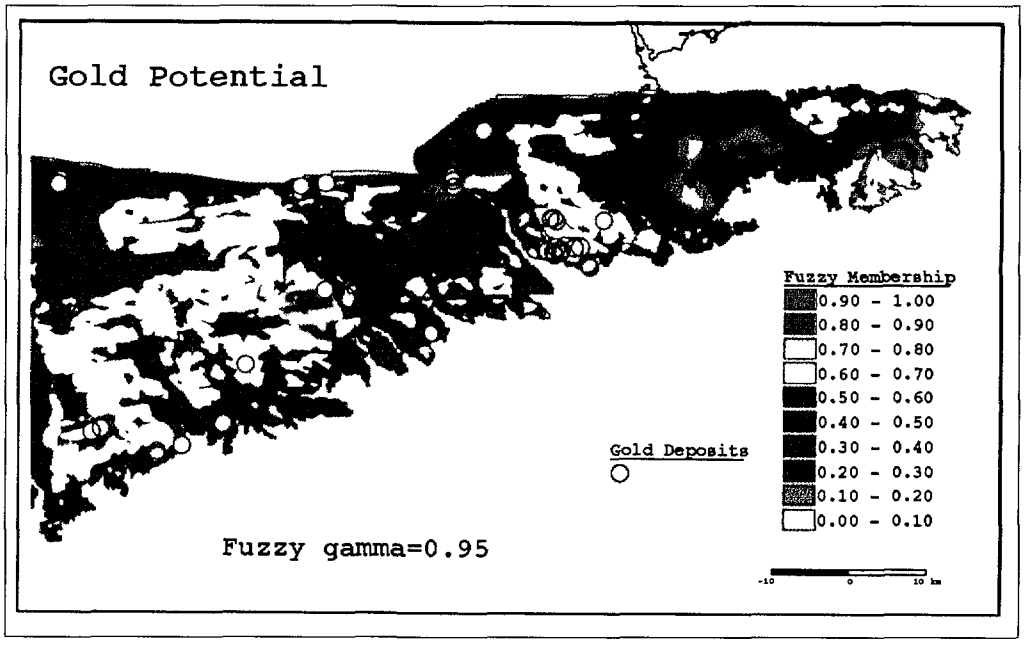



*FIG. 5.15. Fuzzy membership for gold, derived by using the fuzzy gamma operator to combine six multi-class input maps.*

# PLATE 10

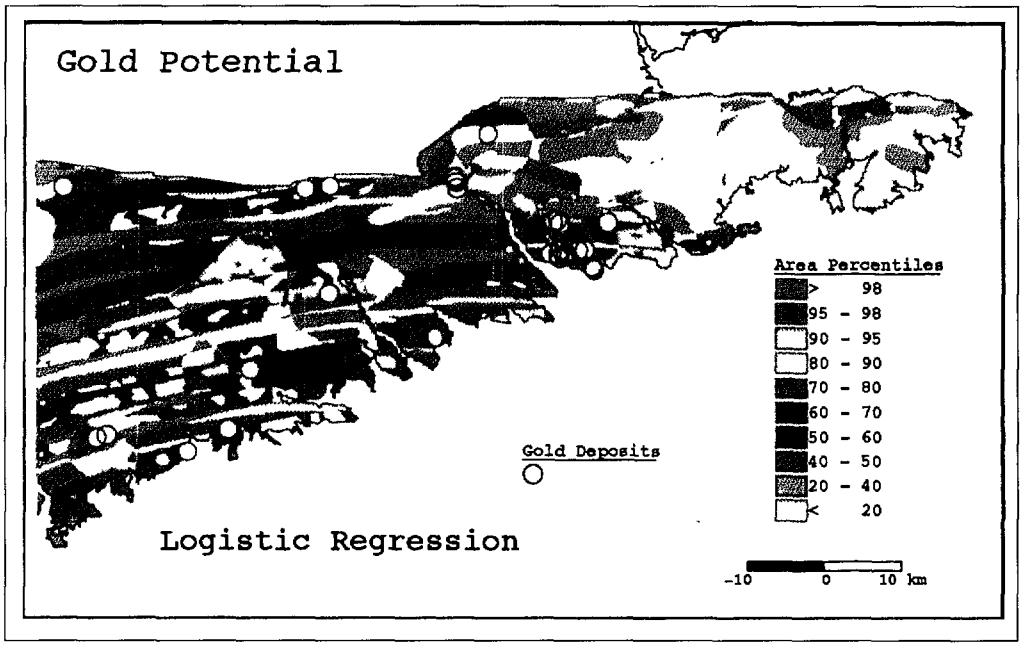



*FIG. 5.16. Gold potential estimated by logistic regression on the same six binary maps used for the weights of evidence prediction.*

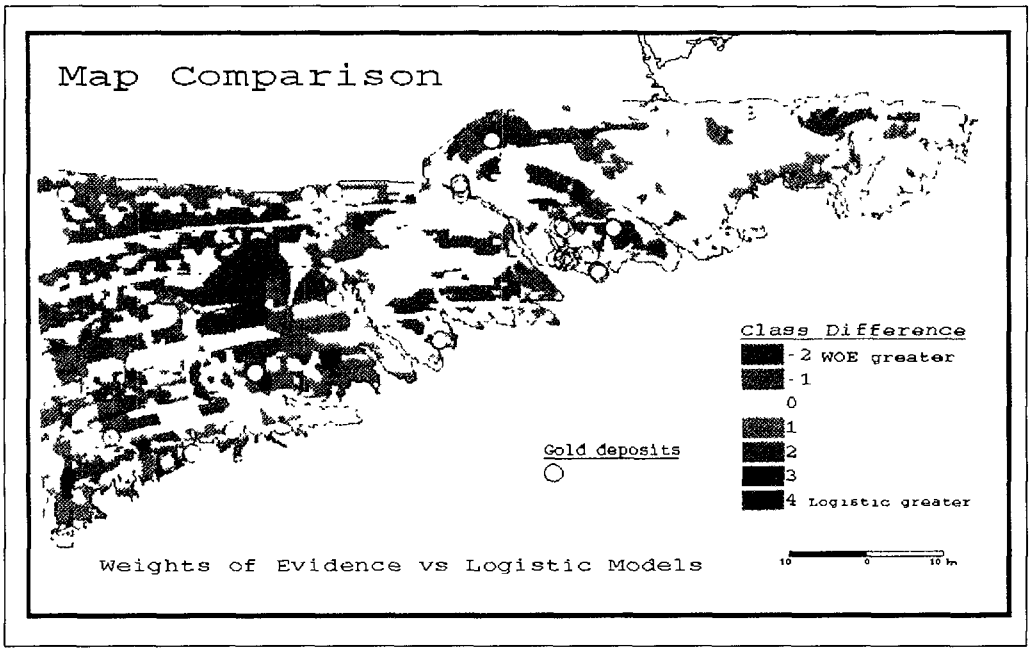



*FIG. 5.17. Comparison of the results determined by weights of evidence and logistic regression. The numbers in the legend are the logistic regression classes minus the weights of evidence classes. The majority of the area is within ± 1 class.*

# PLATE 11

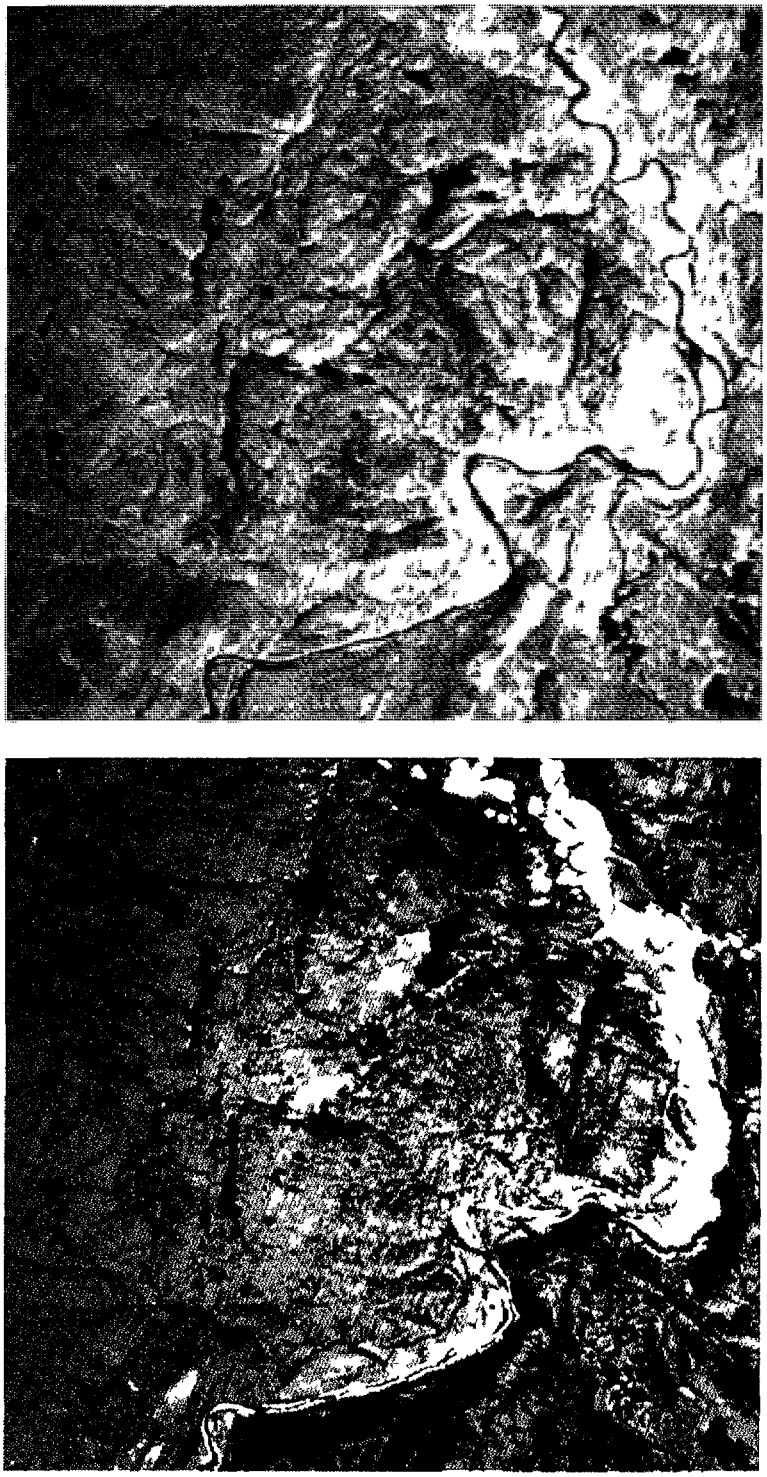


*FIG 7 3 Two colour composite of the Landsat TM image of the study area (a) using TM bands 7, 4 and 3 for red, green and blue, respectively, the three bands have a relatively low pairwise correlation (b) using TM bands 3, 2 and 1, with a high pairwise correlation*

# PLATE 12

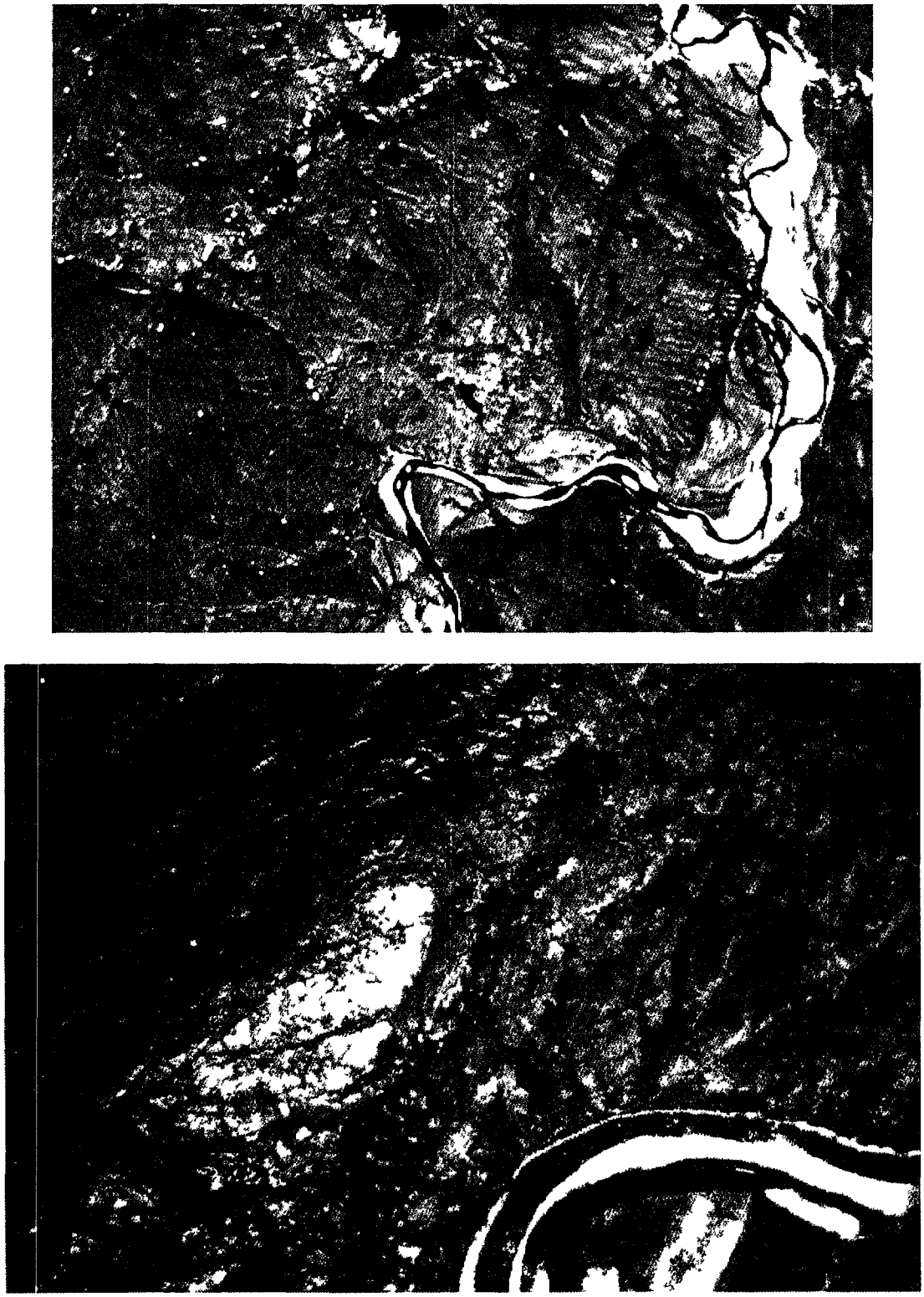


*FIG. 7.4. Ancillary images over the study area: (a) panchromatic SPOT; (b) a scanned aerial photograph.*

# PLATE 13

*FIG 7 5 The digital elevation image obtained from a 1 25 000 topographic map*

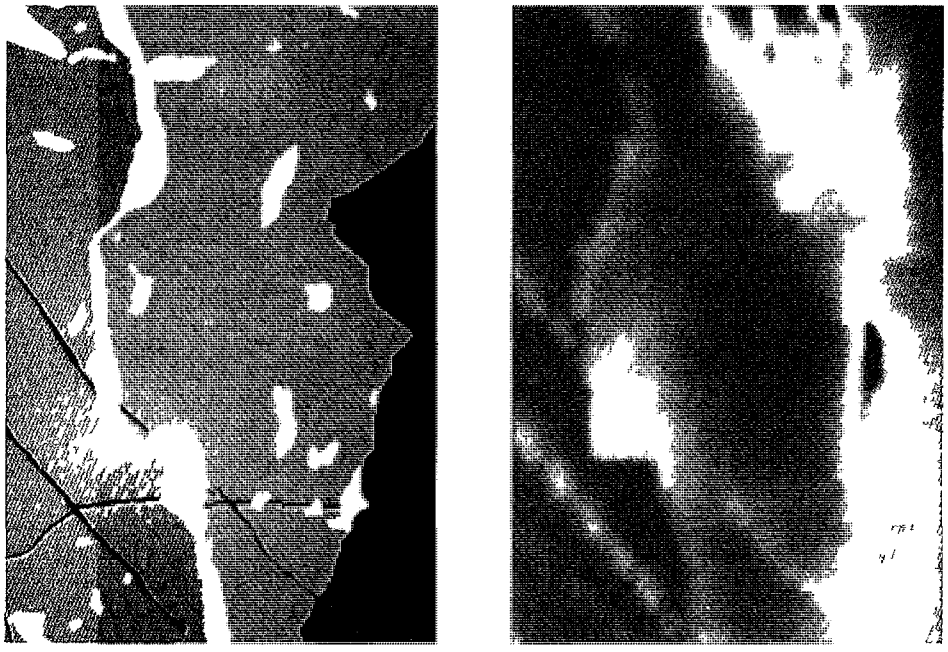


*FIG 7 6 Part of the geological map (a), registered with the detailed aeromagnetic image in (b)*

# PLATE 14

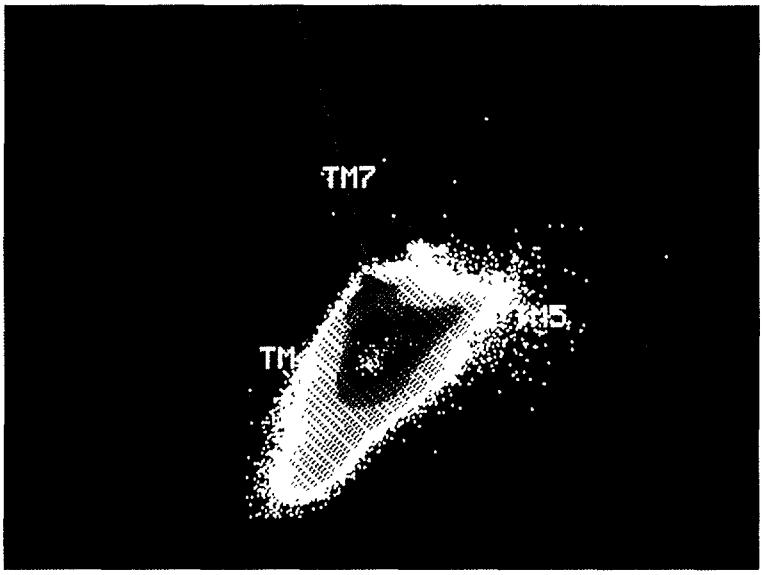



*FIG. 7.17. Landsat TM (4-5-7) feature space with the training data points in the centre.*

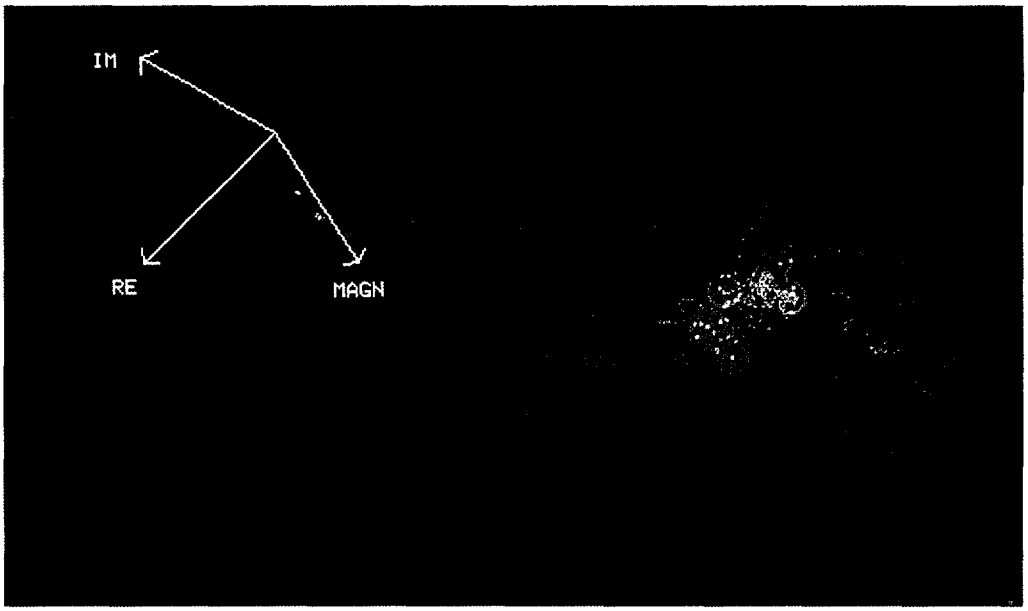



*FIG. 7.18. Feature space of M-EM data (for explanations see Section 7.3.5).*

# PLATE 15

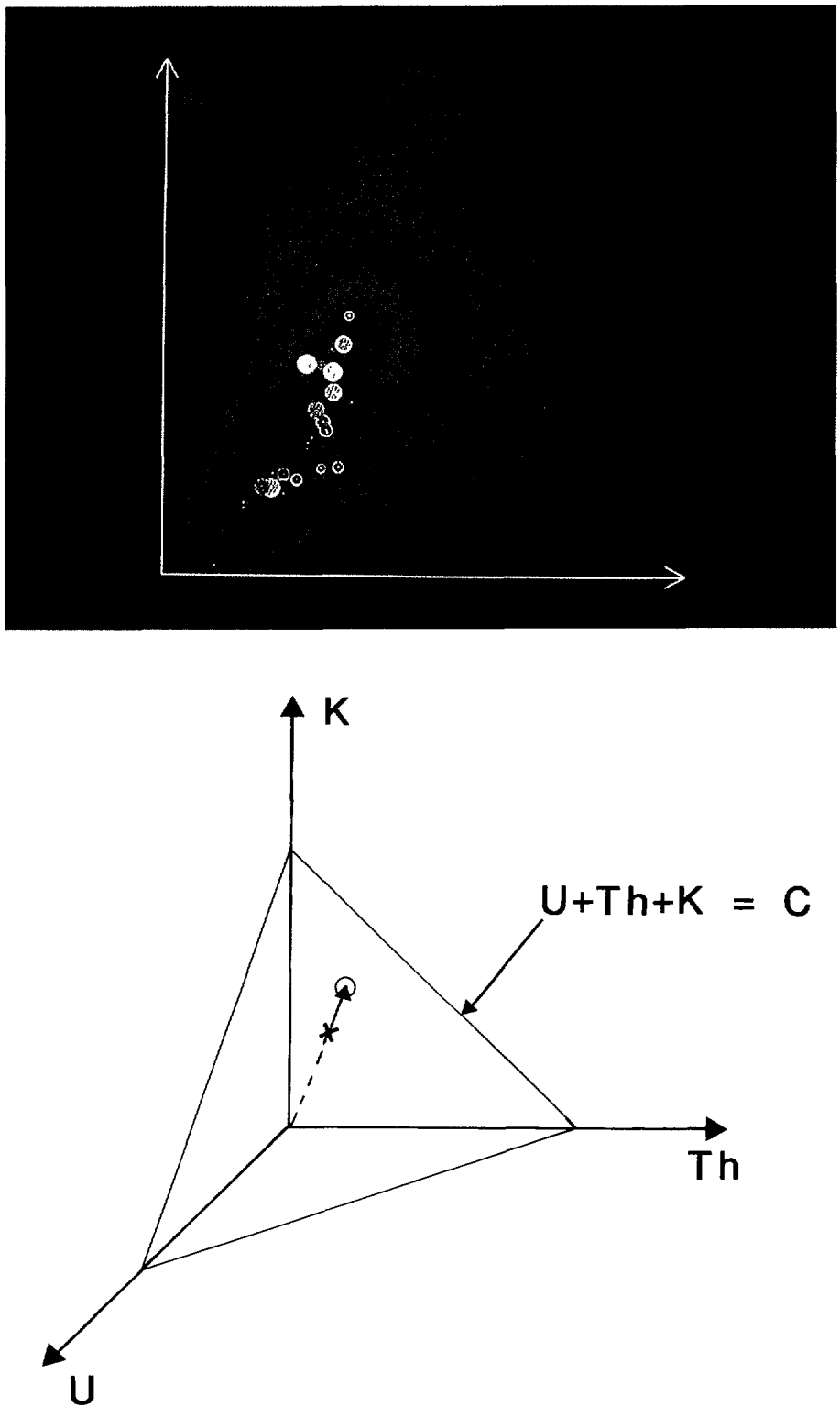



*FIG. 7.19. (a) Two-dimensional feature space of Th and K; (b) ray-projection of the observations onto a plane where the classification is carried out.*

# PLATE 16

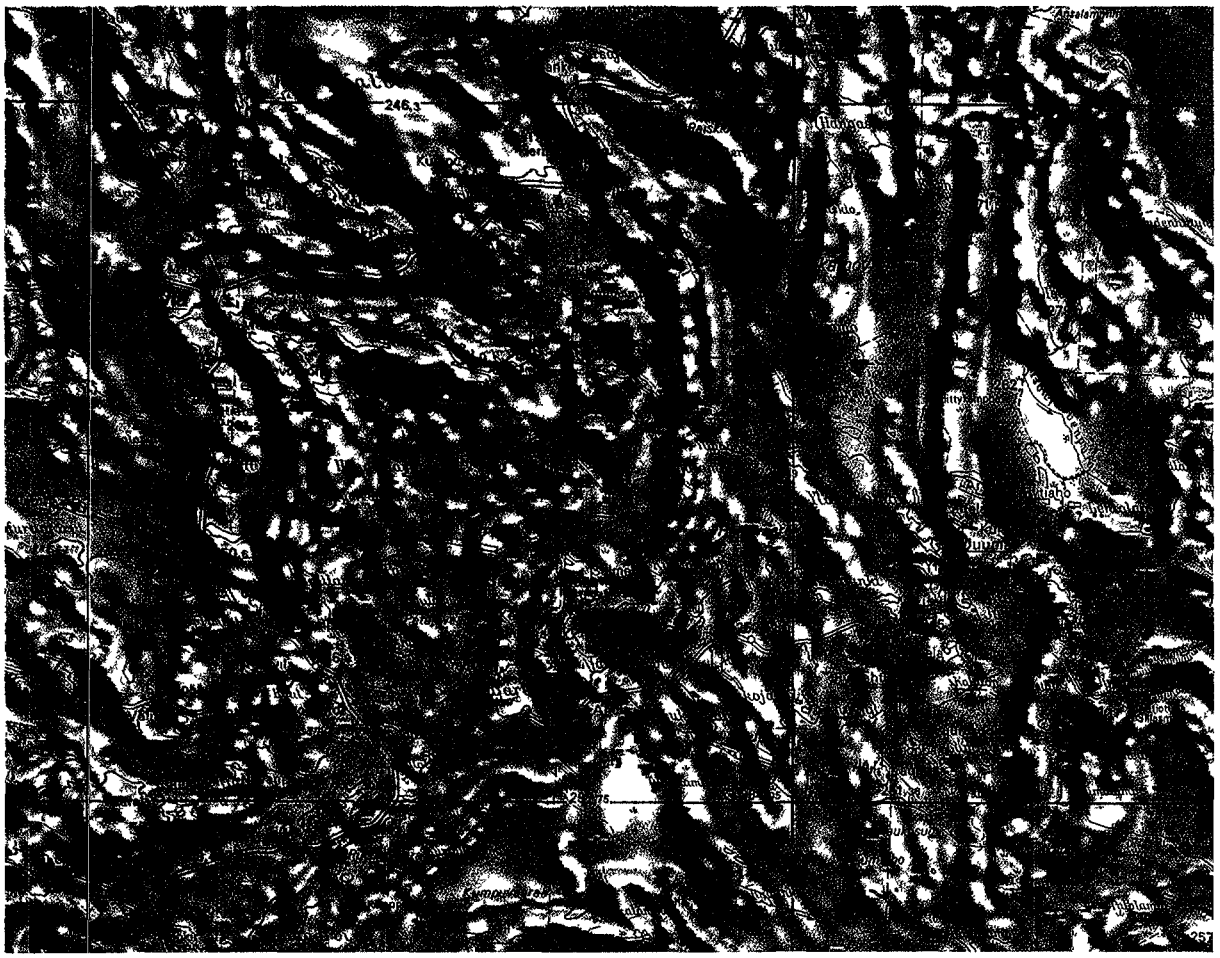


*FIG. 7.20. Analogous sites for gold occurrences on the hill-shaded magnetic map: green = results from classification of Landsat TM features, i.e. vegetation analogies; blue = results from classification of M-EM features, i.e. magnetic and conductivity analogies; red = results from classification of gamma features, i.e. radioactivity analogies. Location of this map is shown in Fig. 7.13. Width of the area is about 20 km.*

94-04665