

United Nations Institute for Training and Research

Explorations in Geographic Information Systems Technology

Volume 5

GIS and Mountain Environments

Edited by

Kristin Schneider and Paul Robbins

Clark Labs

Clark University, Worcester, MA 01610 USA

© 1995, 2001, 2005, 2007, 2009 UNITAR

Palais des Nations

CH-1211 Geneva 10, Switzerland

Introduction

Mountain environments are of special concern to practitioners, designers, and theorists working in Geographic Information Systems. The complexity of these environments stems not only from the geometry of non-linear surfaces, but also from the remarkable diversity of features and facets encountered in mountain ecosystems. For this reason, mountain environments provide both a challenge for the analytic techniques of the computer, as well as for the creative capabilities of the analyst.

This workbook addresses itself to both of these challenges. It is the fifth in the *UNITAR Explorations in Geographic Information Systems Technology* series. Like previous volumes in the series, it highlights not only techniques in GIS analysis, but also problem-solving approaches to exploring and modeling the environment. It similarly incorporates a review paper and a series of GIS exercises relevant to a particular application.

This volume is intended for a wide audience. As an educational tool, it is designed to put a variety of GIS techniques to work on the problem of mountain environments. Although the topics for the workbook originate in a variety of specialized disciplines, including soil science, climatology, ecology, and health services management, the review paper and exercises emphasize GIS methods that may be useful in any number of fields. Common tools and concepts appear throughout the exploration, all being brought to bear on unique problems. As a support for research, the volume shows a number of solutions to problems and questions encountered by practitioners around the world. It may serve as a base for future exploration and as a review of recent applications.

Finally, the work here reflects the complexity and imagination of applications for a field still in its infancy. While some of the research directions included here, like illumination effect mitigation, are quite well-developed, most are not. It is hoped that the volume will serve as the basis for future work in the field, and as a springboard for introducing to a larger audience the questions and problems posed by the use of GIS in mountain environments.

Production of this and other workbooks at the Clark Labs¹ involve a tremendous cooperative effort among many people. We are grateful to Ian Heywood, Martin Price, and James Petch for supplying the review paper. Thanks also to Ian for guiding us through the conceptual stages of the workbook. Much of the data and methodologies for the exercises were contributed by researchers from around the world studying mountain environments. Their generosity and sense of academic cooperation is greatly appreciated. Additional thanks go to Andres Rivera and Tom Millette for their technical assistance.

The workbook team is indebted to Surendra Shrestha, former director of the MENRIS/ICIMOD staff in Kathmandu, Nepal for providing access to numerous researchers who provided much guidance. A special thanks to Govinda Joshi, Rainer Schmidt and Mr. Shrestha.

Kristin Schneider

Paul Robbins

Worcester, 1995

1. IDRISI is a raster-based Geographic Information System for microcomputers developed and distributed on a non-profit basis by Clark University, Graduate School of Geography, Worcester, MA 01610 USA. Development of the IDRISI system has been supported by Clark University, the United Nations Environment Program Global Resource Information Database and the United Nations Institute for Training and Research.

Exercises

The exercises in this workbook require a basic understanding of GIS concepts and methods. Additionally, the exercises are written with steps specifically explained for use with IDRISI, and require some familiarity with that package. Most of the concepts (but not all) are applicable to any raster-based system however, and the explanations are sufficiently general to be applied in other packages.

Because researchers and managers of diverse fields must face the complications of GIS in mountain environments, the exercises presented in this volume are diverse too. They cover a broad range of topics from digital elevation models and interpolation to landscape ecology techniques and complex cost/distance modeling. While it is not absolutely mandatory to follow the order of their presentation, it is helpful; many concepts and techniques explored in detail in earlier exercises may only be treated cursorily in later ones.

The first exercise reviews the technical problems of developing a digital representation of mountain areas and explores the implications of DEM scale on accuracy. It also presents the difficulties in deriving secondary data from elevation models of unknown or questionable accuracy.

Exercise 2 investigates the relationship of topographic variables to land use. Here, changes in land cover are explored in terms of their slope, aspect, and elevation in mountainous terrain.

Exercise 3 explores new methods for examining non-linear distance, highlighting the strengths of a raster environment for modeling movement in mountains. In particular, anisotropic cost distance modeling is used to explore the distances from villages to health care centres in mountainous Nepal.

Exercise 4 introduces the use of GIS to model the relationship of topography to climate, focusing on the use of map algebra to model sun angle, illumination, and insolation potential.

As satellite imagery is a common source of GIS data but can be difficult to interpret in mountainous areas, we include Exercise 5, which explores the effect of topography on satellite imagery. It demonstrates the variability in illumination introduced by mountainous terrain and shows methods for mitigating and modeling this effect.

Exercise 6 is an analysis of the relationship between topographic features and landscape types. Principles from landscape ecology, and a combination of GIS with graduated statistical analysis are used to model the distribution of ecological types in the Pyrenees.

Revision Notes

August 2001

This volume was revised in August 2001 to be compatible for use with Idrisi32 Release 2 GIS and Image Processing software and for distribution in electronic format. The content of the volume is exactly the same as that of the previous version, except that specific instructions for the exercises are given for Idrisi32 Release 2 rather than older versions. In addition, minor editorial revisions have been made. The editor would like to thank Michele Fulk, Kiran Batchu, Takashi Tada, Kelly O'Connor, Laurie Canavan, and Amy Cowgill for their valuable assistance.

Amy Nelson, Revision Editor, August 2001

April 2005

This volume was revised in April 2005 to be compatible for use with the IDRISI Kilimanjaro GIS and Image Processing software and for distribution in electronic format. The content of the volume is exactly the same as that of the previous version, except that specific instructions for the exercises are given for IDRISI Kilimanjaro rather than older versions. In addition, minor editorial revisions have been made.

February 2007

This volume was revised in February 2007 to be compatible for use with the IDRISI Andes GIS and Image Processing software and for distribution in electronic format. The content of the volume is exactly the same as that of the previous version, except that specific instructions for the exercises are given for IDRISI Andes rather than older versions. In addition, minor editorial revisions have been made.

April 2009

This volume was revised in April 2009 to be compatible for use with the IDRISI Taiga GIS and Image Processing software and for distribution in electronic format. The content of the volume is exactly the same as that of the previous version, except that specific instructions for the exercises are given for IDRISI Taiga rather than older versions. In addition, minor editorial revisions have been made.

Mountain Regions and Geographic Information Systems: A Review

D. Ian Heywood¹, Martin F. Price², and James R. Petch²

Introduction

Mountains are distributed across all of the world's continents. They include a vast diversity of environments: from the wettest to the driest; from hot to cold; and from sea-level to 8,848 m at the top of Mount Everest (also known as Sagarmatha and Chomolungma). The cultural diversity found in mountain regions is also great. This is a reflection both of the variety of environments, which permit diverse means of livelihood; and of the importance of mountains as places of refuge, security, and recreation—both spiritual or physical—often at the margins of politically and topographically-defined territories. The diversity, marginality, and strategic importance of mountains—together with vastly different rates of change in different components of their physical, biological, and societal systems—present great challenges for the use of geographic information systems (GIS).

The purpose of this paper is to provide an overview of the current applications, issues, and challenges that are receiving attention by mountain scientists using GIS. Beginning with a consideration of the special characteristics of mountain regions, it examines the extent to which the use of GIS in these regions reflects such characteristics and assesses whether current GIS technology can meet the demands of scientific research and management in mountain areas.

In contrast to the widespread use of GIS in other areas, the technology has received relatively limited use in mountain environments. This must, however, be placed in the broader context of a general naivety towards the science and management of mountain ecosystems. Ives (1992: xiii) for example, suggests that, to the majority of people, mountains have “the appearance of being remote, durable and barely affected by the environmental ills that beset the more densely populated areas of our global living space.” Therefore, funding for the development of appropriate strategies to improve the use and management of these regions is often a low priority. However, in recent years, growing concern over the environmental degradation of mountain ecosystems has meant that mountain issues are gradually finding their way on to environmental and political agendas. An example of this growing interest was the formulation of a Mountain Agenda for the United Nations Conference on Environment and Development (UNCED), held in Rio de Janeiro during June 1992, with the goals of:

- making an authoritative statement on the environmental status and development potential of the world's mountains;
- disseminating this information in the widest possible form; problem; and
- providing some guidelines for a practical response to the problems and challenges of the mountains for consideration by world leaders (Ives, 1992).

All of these broad goals require information, as recognized in Chapter 13, on “managing fragile ecosystems: sustainable mountain development,” of Agenda 21, the programme for action resulting from UNCED (Quarrie, 1992). Yet, at the present time, the necessary information either does not exist or is far from comprehensive in its coverage. In this respect, the Mountain Agenda and the consequent Chapter 13 of Agenda 21 can be seen as an excellent basis for arguing for the wider use of GIS to assist in developing our understanding and improving our approaches to the management of moun-

1. Department of Environmental and Geographical Sciences, the Manchester Metropolitan University, John Dalton Building, Chester Street, Manchester M1 5GD.

2. Environmental Change Unit, Oxford University, 1a Mansfield Road, OX1 3TB.

tain environments.

Several organisations at regional, national, and global scale have already recognized this potential and are developing long-term monitoring programmes in which GIS will play a central role (Halpin, 1993, Koshkariov et al., 1993, Walsh et al., 1993). In all cases, the objectives are to record baseline information about the state of mountain regions and the anthropogenic stresses that affect them; to evaluate methods for the continued monitoring and sustainable development of these environments; and, in some cases, to evaluate future scenarios deriving from the interaction of biophysical and societal processes, usually through simulation models (Brzezicki et al., 1993, Halpin, 1993, Schaller, 1993, Vasconcelos et al., 1993).

It needs stating from the outset that there is nothing unique about the character of GIS application in mountain areas. This is to be expected, since although we can regard mountains as distinct areas of the earth's surface with a common set of issues, they are not unique. In fact, these issues – including resource assessment, hazard prediction, environmental impact assessment, and ecosystem management – are determined not only by the nature of any particular landscape but by society's collective objectives. Nevertheless, the use of GIS in mountains requires some special considerations which derive both from the particular characteristics of mountain environments and the peculiarities of data collected in these complex, dynamic regions, and also from the limitations, originating in current GIS software and ecological theory, on our ideas about how mountain systems work.

The great variability of mountain environments at all spatial and temporal scales requires great care, both in the choice of scales for data collection and storage, and in extrapolations from locations with high-quality, long-term data to data-poor areas – even over quite small distances. Methods of assessment and extrapolation developed in lowland environments may be quite unsuitable for heterogeneous, fast-changing, mountain environments; when applied in these areas, the assumptions embedded in such methods must be made explicit and assessed for their applicability for a specific task, using detailed local knowledge. Further problems include:

- modelling extensive variations in relief and the shape of terrain with a technology that essentially remains two-dimensional;
- ground-truthing remotely-sensed, or model-generated, data in isolated environments;
- handling the temporal aspects of rapid environmental change; and
- communicating the results of GIS-based analyses to local populations hampered by lack of understanding of such technologies.

The Special Characteristics and Complexity of Mountain Areas

Mountains and uplands comprise about one-fifth of the world's terrestrial surface, and are directly or indirectly important for more than half of the world's population (Ives, 1992). They are home to about one-tenth of the world's population; supply natural resources – food, wood, and minerals – to an even greater proportion of the global population; are at the upper end of most of the world's river catchments, providing water, nutrients, and energy to those living both nearby and at distant locations downstream; provide environments for recreation and tourism for visitors from both nearby and far away; include centres of biodiversity and refugia for relict species and communities; and are of great spiritual and aesthetic significance to many people.

Although there are numerous definitions of what mountains are, there is general agreement that they are areas of high relief. They range from isolated peaks and islands to complex systems of considerable extent (Gerrard, 1990). Thus, the physical characteristic that best defines mountains is their three-dimensionality, which produces contrasting environments at different elevations. Superimposed on this altitudinal zonation, however, are variations that derive from the aspect, slope, and topography of a particular mountain or region (Barry, 1992b). In fact, it is this three-dimensionality that poses the greatest challenge for modeling these regions using GIS, for the simple reason that most GIS and the data they incorporate still treat the world as if it were flat. Nevertheless, GIS used in mountain areas usually incorporate digital terrain models, which permit the representation of the three-dimensional nature of mountains (Stocks and Heywood, 1993).

The importance of the third dimension can be summarised by a consideration of global and local climate. In the northern hemisphere, south-facing mountain slopes receive more radiation than north-facing slopes. Consequently, the former tend to have a longer snow-free period, larger ranges of temperatures at diurnal and annual scales, and drier micro-climates. The degree of these differences also varies with slope angle and latitude. The topography of mountain regions has further effects on local and micro-climates both in valleys and on upper slopes and summits. In valleys, cold air collects during stable meteorological conditions, usually experienced in winter. In such “inversion” conditions, the typical decrease of temperature with increasing altitude does not apply: upper slopes are far warmer than valley bottoms and lower slopes. However, upper slopes and summits tend to be especially harsh environments, characterized by strong winds, low temperatures, and moisture deficits. Consequently, a GIS is valuable because, despite its two-dimensional view of the world, it is still possible to model such features as aspect, slope, and height.

The ability of a GIS to model the nature of the terrain is essential because the local and micro-climatic variations described above strongly influence the biophysical components of mountain environments: air, water in its different phases, soils, vegetation, and fauna. Precipitation and water storage and flow all vary with altitude and aspect. Thus, for example, in the northern hemisphere, mountains may have remnant glaciers and permanent snowbeds on their north-facing slopes, but none on their south-facing slopes. Both soils and vegetation tend to occur in “belts” at different altitudes, their concentricity around a mountain or range offset by the variations deriving from differences in local and micro-climate which result from aspect; for instance, the tree-line is typically lower on north slopes in the northern hemisphere.

In addition, variations in local and micro-climates also influence patterns of housing, agriculture, and recreation. Valley floors are often particularly desirable locations for settlement and agriculture. However, the stagnation of cold air, which can result in the concentration of airborne pollutants from sources as diverse as wood-fires and automobile exhausts, means that winter resorts are often developed on upper slopes. These may replace small high settlements, previously used only by herdsmen bringing animals to pastures during the brief summer season. This pattern of resource use, which permits domesticated animals to benefit from a range of food sources in different seasons, is found in many mountain regions.

To develop an understanding and appreciation of the optimal locations for settlements, it is also essential that the GIS should be able to identify areas affected by the likelihood of “natural” hazards – including avalanches, rockslides, floods, and forest fires – whose distribution is influenced by complex interactions between local climates, human activities, and soil, bedrock, and vegetation characteristics (Hewitt, 1992). While likely occurrences of some of these hazards are somewhat predictable, those of others – such as mass movements triggered by earthquakes and volcanic eruptions – are less so, although they may have far more extreme effects than more frequent events. This is not simply a mapping problem and a GIS must be capable of addressing the issue of how, when, and where these high-risk events are likely to occur.

In the national and global context, a further important characteristic of mountain environments is that they tend to be marginal areas which are physically and/ or culturally distant from centres of political power and directly contribute little to national economies. Mountain agriculture and forestry are often seen as a drain on national and even regional budgets although they provide both direct and indirect benefits for considerable numbers of people. In general, per capita levels of investment tend to be far lower in mountain areas than in adjacent lowlands, and investments (e.g., large hydroelectric schemes and tourism developments) may not directly benefit long-established mountain dwellers, who may even be disadvantaged or displaced by such projects. Placed in a GIS context, the marginality of mountain environments means that there is a need for a tool to assist policy makers, planners and environmentalists at local, regional, and national levels to develop strategies for the economic and ecological management of these regions.

The geographical and economic marginality of mountain areas is often exacerbated by their political sensitivity. Many mountains form natural barriers, so that national boundaries (often in dispute at the present or in the past) run through them. Thus, the sensitive nature of these border areas – and, not infrequently, armed conflict – may greatly hinder the collection of essential data through air- or space- borne remote sensing and field work (Schweinfurth, 1992).

Developing and Using GIS in Mountain Areas: Data Issues

Potential and actual users of GIS in mountain environments are confronted by a wide range of issues with regard to the

data that can be incorporated and used. The availability and collection of data may be affected by historical, political, climatological, topographic, and many other considerations. Equally, special attention must be given to the processing, analysis, and modelling of data relating to mountain environments because of their spatial and temporal complexity. These issues are discussed below.

Regional Diversity

Superimposed on the great physical and biological heterogeneity of mountain regions are diverse histories and patterns of human use. When developing a GIS, this is important since the patterns and histories of human use of mountain regions greatly affect the types and availability of existing data and may restrict the use which can be made of them. A useful typology in this respect is provided by Grötzbach (1988), who differentiates between “young” and “old” mountains. The former are the relatively sparsely-settled mountains colonized by the Europeans in recent centuries. Examples include the mountains of North America, Australia, and New Zealand. Here, perhaps because of the introduction of European concepts of property rights and a market-oriented approach to development, there is very often a wealth of geo-referenced information suitable for use in GIS. It usually exists in the form of paper maps, land inventory surveys, and property ownership records. More recently, the application of remote sensing has led to the development of more regular broad-based land inventories suitable for use in GIS.

Most of the “old” mountain regions are relatively densely settled. They can be divided into three categories. The first of these is characterised by a decline in traditional agriculture and forestry, linked to depopulation except in areas experiencing a growth in tourism, which in many cases has become the basis for the economy. These regions are typical of many of the mountains of western Europe. The long human history of these regions means that they too are often data-rich in terms of information suitable for inclusion in a GIS. Spatial information is often available at a much more detailed scale than for young mountain regions, and the temporal range is generally much greater.

The second category of old mountains includes most in less-developed nations. Traditional subsistence and/or herding agriculture is the dominant land use and there is a tendency towards over-population. Most of these regions are data-poor, with the exception of those which found themselves subject to colonialism and where detailed mapping was undertaken. Thus, the majority of available spatial information tends to have been derived from satellite surveys in recent years; it is very often owned by the military. The political instability of many of these regions makes access to data particularly problematic.

The third category of old mountains has, until recently, largely been characterised by collectivised or nationalized agriculture and forestry. These include the Carpathians and the mountains of the former Soviet Union and China. With the fall of the communist regimes in all of these regions except the latter, their social, economic and political structures are now in a state of flux, as are the availability of, and access to, geo-referenced information. While it is very hard to make any general statements, two examples may be given. First, work on GIS in the Altai mountains of Siberia (Carver et al., 1993) has revealed a plethora of geographic information that is as extensive in its coverage as that for many of the young mountain regions. Second, the Atlas of the Tatra National Park in Poland (Trafas, 1985), with a remarkable array of thematic maps, many of them at scales of 1:25,000 or better, is probably unmatched even in western Europe. Both of these comprehensive data sources are ripe for inclusion in GIS. Nevertheless, in general, scientists in this third group of old mountain regions not only suffer from a lack of hardware, software, and training, but often face the additional problem that existing paper maps often contain intentional distortion and misleading generalisation of key geographic features such as roads, towns, and rivers, most probably carried out for “security” reasons.

Data Sources

Three main sources of data have been used in the development of GIS for mountain areas: maps, field survey data, and remotely-sensed imagery from aerial photography or satellites. Such imagery is typically used to provide a basic classification and inventory of land use/ cover types, as described, for instance, by Parrachini and Folving (1993) and Walsh et al. (1993). However, the availability of such classifications varies greatly between mountain regions. Aerial photography may not be available or possible, particularly in areas which are politically sensitive; and digital remote sensing data may be of insufficient spatial and/or temporal resolution. In addition, mountain areas have strong topographical gradients which

markedly affect reflectance patterns; and the classification of large areas is often hindered by frequent and widespread cloud cover.

For any combination of these reasons, full coverage of an area may require the collection of images over many days or even seasons. This may only be a partial solution since, because of the rapid phenology of many mountain plant communities and crops, images of an area obtained only a few days apart may have significantly different reflectance characteristics. Time of day is also a critical constraint: even at solar noon, shadows on steep slopes can make classification very difficult. Finally, short snow-free seasons limit the period during which useful imagery can be obtained. While air- or space-borne remote sensing can be of great value in developing land, cover/use classifications of mountain areas, field work remains essential and this, again, may be hindered by political considerations.

Enhancement and classification of satellite images requires the control of topographic effects. This is best achieved using a GIS to isolate these effects of aspect and gradient (Parrachini and Folving, 1993). Global Positioning Systems (GPS) have recently been used with GIS to detect sites in mountain areas which are used in image classification (Rodcay, 1991), and can be used to determine sampling positions and heights to create grids for photo rectification (Haefner and Hugentobler, 1985). Exercise Five in this volume presents two accessible approaches for managing topographic effects in satellite imagery. Such enhanced images have been used for many applications of GIS in mountain regions, including resource evaluation (Walsh et al., 1990), land cover mapping (Downey et al., 1990), biodiversity analysis (Rodcay, 1991), and assessment of forest fire effects (Butler et al., 1990).

Data Models for Mountain Environments

One necessity in the development of a GIS for a mountain area is the ability to model its three-dimensional complexity. From the literature, two broad approaches can be recognized. The first involves the use of digital terrain models (DTMs), also known as digital elevation models (DEMs), to provide a categorization of zones or elements of a mountain area according to slope, elevation, and aspect. The second uses landscape units, constructed from a synthesis of environmental data, which reflect the character (structure, sustainability, and responsiveness) of an area rather than its physical form (height, shape, and exposure) alone.

The DTM Approach

The DTM provides a framework for the use of GIS in mountain areas by capturing the three-dimensionality of mountain environments as a two-dimensional mosaic. This can be constructed of either regular or irregular tessellations, using the grid or triangular irregular network (TIN) methods, respectively. For each tessellation, the three major terrain elements of height, aspect, and slope angle can be recorded as attributes (Weibel and Heller, 1991). The spatial relationships between these attributes are then used to portray the three-dimensional character of the area. Given the critical importance of terrain in understanding mountain environments, DTMs are an essential element of GIS for mountain areas. Exercise One of this workbook explores issues critical to the construction of DTMs.

The Landscape Approach

The landscape approach does not deal directly with the problem of capturing the physical shape of terrain. Instead, it incorporates higher levels of knowledge about the character and behavior of mountain systems. The approach draws on the principles of landscape ecology (e.g., Forman and Gordon, 1986). As a first step, a landscape model of the unchanging physical characteristics of an area is constructed from ground survey information, aerial photography, and base maps. These include relief maps, from which facets of the landscape are defined, each with a specific combination of altitude, aspect, and slope. In practice, these maps are drawn by hand, and include a limited number of categories decided by the cartographer. Other data on geology, soil types, hydro-meteorological conditions, and vegetation are used to produce a typology of the landscape units that represent permanent conditions of the landscape (Bucek and Lacina, 1981).

While the development of such "landscape maps" has, to date, been time-consuming because of the need for expert knowledge, new methods using knowledge-based systems are being pioneered (Burkmar et al., 1992). These landscape maps are then used in conjunction with land cover or land use maps, or satellite remote-sensing products (Downey et al.,

1991), to assess the appropriateness or stability of use in each land parcel in relation to the natural conditions. Such assessments have two parts, considering both the suitability of a site and its position in the two-dimensional structure of stable and unstable sites (Petch and Kolejka, 1993). One basic problem of the landscape approach is that assessments of stability, generated by the overlay of landscape units, are not easily interpreted by “non-experts.” This presents particular problems in the use of this information in policy-making. However, the landscape approach does represent an alternative method of incorporating knowledge of the complex nature of mountain areas into GIS. See Exercise Six of this workbook for an example of the landscape approach applied to ecological modelling.

Data Quality and GIS Techniques

Many of the problems hindering the application of GIS in mountain environments relate to the quality of the available data and their suitability for use in GIS. Data for almost all aspects of mountain ecosystems are very heterogeneous in their length and frequency of record, spatial coverage, and availability. Three examples of data issues, each with an increasing degree of challenge posed by the problems deriving from heterogeneity, are discussed below: overlay techniques, process modelling, and extrapolation.

Overlay techniques are perhaps the most frequently used to obtain additional information from existing data sets. The problems engendered by overlaying maps of different scales and/or based on various coordinate systems are well-known in all environments. However, the three-dimensional complexity of mountains can exacerbate these problems to an extreme extent. One example is provided by Beissmann (1989) who attempted to create a GIS of the Wallackhaus area of the Hohn Tauern National Park, Austria, using a number of high-quality thematic maps, all based on detailed field research in the 1970s and earlier. However, they did not have a consistent, accurate cartographic base and were at different scales. As a result, there was considerable disagreement in the location of land-cover categories; even those that should have been well-defined and consistent between maps, such as talus slopes. Such problems may become less critical as more spatially-consistent technologies become available. However, great care must be taken to use these effectively; and both fundamental issues and time-consuming problems will still remain when using data sources that are uncorrected or have different scales, resolutions, and degrees of generalization.

With respect to the modelling of environmental processes, it is essential to correctly choose both the appropriate variables and the temporal and spatial scales at which they are defined. An example is provided by Wadge (1988), who has shown how the accuracy and quality of a single data set, in his case a DTM, can influence the results of slope process models. The research showed that most DTMs are inappropriate for determining the dynamics of slope stability where the value of every grid cell is crucial to the model. Similarly, dynamic slope-flow models fuelled with DTM data of the wrong spatial resolution may lead to a magnification of the scale of the problem, so that material is sent down the incorrect path. Nevertheless, models of slope energy balance appear to be less sensitive to individual pixel values, so that it is quite plausible that DTMs derived from topographic maps at scales greater than 1:1,000 may provide an overview suitable for the development of zone mapping. In general, the choice of (spatial and temporal) scale of the data is critical for successful, reliable process modelling; data whose resolution is either too great or too small may provide results that are incorrect, misleading, or even meaningless --- and often untestable.

Extrapolation is an activity for which GIS are being widely promoted and often used. As noted by Halpin (1993) this can be very risky without a deep understanding of the inter-relationships of the variables involved in the extrapolation, and the risks of inaccurate or improbably extrapolation increase with the number of variables and the complexity of the environment. As Haslett (1993) has recently shown using fractal analysis of the Berchtesgaden GIS (Ashdown and Schaller, 1990), mountain environments are the most spatially complex. Thus, for any phenomenon, extrapolation can be highly questionable, even between locations with reliable data.

Mountain climates provide many examples of complex, dynamic phenomena (Barry, 1992a). They are characterized by marked diurnal and seasonal cycles, with high variability at all spatial scales. As mentioned above, climate is a critical controlling factor in the distribution of environmental components and the uses of mountain environments, so that the possibility to extrapolate – in both time and space – from existing information is highly desirable and can theoretically be achieved using GIS. Yet, given the great variability of climates at all temporal and spatial scales, the useful characterization of the climate of any mountain area requires long-term records from dense networks of stations at a wide range of alti-

tudes and on slopes of different aspects.

Unfortunately, such records are rare. Climatic data tend to be more readily available from settlements close to the mountains or in valleys, rather than from mountain sides and summits. The length of records is often inadequate to provide a realistic assessment of the mean and, especially, the extreme values of relevance to both human beings and other components of mountain ecosystems. This problem is exacerbated by changes in the location of recording stations – a critical problem in areas with large variation over small distances – or because data collection is only seasonal; for instance, in winter to provide information for predicting avalanches or spring run-off.

The conclusion with respect to this and the preceeding data issues is that the use of GIS must involve awareness of the limitations of not only the available data but also the understanding of environmental processes and the technology in use. Given such awareness, GIS can be a valuable technology for descriptive, analytical, and evaluative purposes.

GIS Applications: From Inventory to Management

GIS have been used for many objectives in mountain areas around the world. To all intents and purposes, these applications reflect environmental, cultural, and economic issues that have come to the fore in recent years including, for example, forestry, mining ecology, tourism, hazard mapping, visual impact assessment, and climate change modelling. These broad application areas have seen the use of GIS in two main ways: first, as tools to assist in resource inventory and the integration of data and, second, as a mechanism for analysis, modelling, and forecasting to support decision-making (McKendry and Eastman, 1990). In many cases, the GIS is used in both capacities. The following sections present four categories of GIS applications: regional resource inventory and planning; evaluation of natural hazards; research and resource management in and around protected areas and; simulation and prediction. However, the following sections should not be construed as an exhaustive review of a rapidly-evolving field whose literature is widely scattered and often rather inaccessible.

Regional Resource Inventory and Planning

Data inventory and integration has been the most consistent use of GIS since the development of the first GIS for mountain areas – such as the Lluçanes area of the Spanish Pyrenees (Alegre, 1982) and the Grindelwald area of the Swiss Alps (Steiner and Zamani, 1984) – on mainframe computers. Most mountain GIS have been developed with data at a limited range of scales for specific purposes in well-defined areas. While a few attempts have been made to create multi-purpose data banks or data bases for mountain areas, there have been considerable difficulties in designing such versatile GIS. One example was the creation of a GIS for Nepal, through a joint project involving the Global Resource Information Database (GRID) of the United Nations Environment Programme (UNEP), in collaboration with the International Centre for Integrated Mountain Development (ICIMOD), in Kathmandu (Simonett, 1993). The objectives of this project include the creation of environmental data bases at different scales and resolutions: national (1:1,000,000 – 1:3,000,000), zone district (1:50,000 – 1:500,000), and local (1:10,000 – 1:50,000). In fact, the local and zonal GIS are partial in coverage and oriented towards particular objectives. There is, nevertheless, a general attempt to approach the problem of a central GIS resource. Exercises Four and Five utilize data from this project.

One example of a smaller scale of operation is the design of a multi-user GIS for the Zdar Hills protected area of the Czech Republic (Pauknerova et al., 1992). This is a GIS with over 20 data layers, intended for use by different groups and agencies (Downey et al., 1990). The design of the system includes two specific elements which permit its use for particular purposes: a system of land units, and a territorial system of ecological stability. Using ideas of landscape ecological management (Michal, 1992), these provide the methodological basis for using the same data layers for different purposes. The GIS can therefore be used as a general tool for data integration. Similar work has been done in the Malcantone region (65 km²) of Switzerland, in an intensive study of landscape change and landscape protection (Haefner and Hugentobler, 1988; Haefner et al., 1991). In this project, the analysis of landscape, using remote sensing and GIS, is the basis for studying economic and environmental problems, such as conflicts between land uses with respect to landscape protection, and the identification of areas at risk from flooding, in order to develop appropriate land use planning and management strategies.

Schaller (1993) describes the conceptual bases for a very detailed GIS for the management of Berchtesgaden National

Park and the neighbouring region in Bavaria, Germany. The paper is unusual in that it presents the ecological theories on which the development of the GIS was based; most GIS are initially created as tools for data integration and analysis without a theoretical framework. The approach developed in Berchtesgaden has since been applied to other mountain areas in China, Kenya, and New Zealand (Ashdown and Schaller, 1990). Parrachini and Folving (1993) present the initial development of a GIS for regional planning and modelling in the Italian Alps. The principal objective of the GIS will be to provide low-cost, accurate mapping and methods for monitoring environmental conditions, with particular regard to hydrological resources.

A number of GIS studies in their early phases are described by Koshkarirov et al. (1993) at scales from the entire Commonwealth of Independent States to individual rayons (districts). At present, the main aim of this work is the generation of digital atlases which, in the future, will be used for regional planning. Pulsford and Ferrier (1993) describe the implementation of a GIS to support the multiple responsibilities of the National Parks and Wildlife Service (NPWS) of New South Wales, Australia, both in National Parks and throughout the state. The approach is notable for its decentralised approach and the use of low-cost equipment. The GIS has been used to assist in a wide range of management and research activities.

Within more narrowly defined objectives, GIS have been created as data integration tools for many mountain areas. Among the objectives of a project on forest resources in Nepal were the documentation of historic change, the determination of relations between use and site factors, and the prediction of future resource use (Schmidt and Schreier, 1991). Exercise Two in this volume is an early example of their work. GIS appears to be playing a key role in wider strategies of resource and environmental management in Nepal (Schreier et al., 1990, 1993). A similar GIS strategy has been developed for the Great Smoky Mountains National Park, USA (Parker et al., 1990), where the GIS is used to analyze the spread of forest diseases and to produce inventories of forest types and biodiversity.

One of the significant potentials of GIS is to use their three-dimensional and graphic capabilities to make inventories of not only the usual biophysical and cultural resources, but also visual resources. Thus, current landscapes can be portrayed and the effects of future planning scenarios assessed. Culbertson et al. (1993) present a number of case studies in which visual resources were considered as an integral component of the regional resource base when developing plans for the future of mountain areas – with recreational facilities, settlements, transportation corridors and, in two cases, mines – in Brazil, Canada, and Japan. The methodology proved valuable both as a means for integrating a wide range of information and also as a basis for public involvement in the planning process. Their final example, from Colorado, shows the potential of GIS as an integrative tool for planning the use of multi-jurisdictional resources. Miller et al. (1993) also focus on the assessment of recreational resources through the evaluation of scenery and the visual impacts of land management practices, particularly forestry, in the Cairngorm Mountains of Scotland.

As shown by some of the examples above, the greatest and most interesting challenge for mountain scientist is the use of GIS technology in an investigative role: seeing answers to “what-if” questions such as: which areas are at risk from landslides?; what will be the economic and cultural implications of particular tourism/ agricultural activities?; how will deforestation affect soil erosion?; or, what are the potential implications of long-term environmental change? To answer these questions, GIS can provide a wide variety of analytical, modelling, and forecasting modes. While resource inventory is the essential first step, the greatest potential of GIS technology is its use to link resource data with explanatory and predictive models and expert knowledge. The following sections consider three general themes – natural hazard evaluation, research and resource management for conservation, and simulation and prediction – recognizing that any GIS may have the potential to be used for any or all of these purposes and for many others.

Evaluation of Natural Hazards

Given the dynamic nature of mountain environments, natural hazards – such as avalanches, landslides, fires, and floods – are major topics of concern for management. GIS can be used in many ways to increase our understanding of these phenomena and assist in developing adaptation and prevention strategies. Van Westen (1993) provides an overview of the use of GIS for landslide hazard zonation, using examples from the Andes of Colombia. A range of analytical approaches are presented and evaluated with reference to their applicability with respect to different levels of data availability and scales of analysis. In contrast to this classical approach – based on the integration of remotely-sensed imagery, maps, field work,

and expert knowledge – Vasconcelos et al. (1993) present an approach to fire spread modelling in which a GIS is linked to a knowledge-based simulation model. This work is at an early stage of development, but shows considerable promise for developing predictive understanding of discontinuous processes of which fires are but one example in mountain environments.

The use of GIS to help in modelling the dynamic nature of mountain ecosystems is also characterised by the work of Walsh et al. (1993) who have been involved in producing a probability map of where, and on what terrain, avalanches are likely to occur in Glacier National Park, USA. Their multi-criteria approach weights the importance of different data layers (height, aspect, and slope angle) within the GIS relative to their importance in the avalanche formation process. The study is a clear example of how the GIS may provide valuable information for the modelling of physical processes. Both Walsh et al. (1993) and Pulsford and Ferrier (1993) also consider aspects of fire management, from the analysis of past events to potential assessment and suppression. Comparable work has been undertaken by van Wageningen et al. (1990) in Yosemite National Park, USA.

Research and Resource Management in and around Protected Areas

GIS has been widely used as a tool to aid the design, management, and monitoring of National Parks and other protected areas (Willison et al., 1992). Numerous case studies reveal the benefit of applying this technology for both policy and research.

One of the major programmes of research in a National Park began in the early 1980s in Berchtesgaden National Park, Germany (Schaller, 1993). Part of this programme was the development of a zoological GIS which includes topographical data, habitat information derived from literature reviews and field work, and the results of field surveys (d'Oleire-Oltmanns and Franz, 1991). The system has been used for habitat analysis and space distribution models of both individual species (d'Oleire-Oltmanns, 1987; Schuster, 1990) and species assemblages of vertebrates and invertebrates (d'Oleire-Oltmanns, 1990; Haslett, 1990) and for simulating changes in populations (d'Oleire-Oltmanns et al., 1991). The visual outputs which derive from this work are used for both research and resource management. A GIS has also been used for zoological research in Langtang National Park, China, where the locations of red pandas, recorded using GPS technology, have been compared with habitat evaluations for different seasons. The resulting maps have been used to estimate populations and develop management plans (Yozon et al., 1991).

Another theme of research has been the monitoring of landscape change. One major study considered the twelve National Parks of England and Wales, most of which are in upland or mountain areas. The study, based on multitemporal aerial photography and field surveys, showed how GIS could assist in developing a methodology for time-series analysis of complex data sets (Bird and Taylor, 1990). Satellite remote sensing images have also been used as a multi-temporal data source, combined with existing base maps and maps deriving from field research, for determining management regimes in Mercantour National Park, France (Claudin et al., 1993).

Walsh et al. (1993) present a summary of a major ongoing research programme conducted over many years in Glacier National Park, USA. Building on considerable previous research in the Park, the programme has included four stages. Beginning with the creation of an integrated database from a wide variety of sources, the work proceeded to the evaluation of various landscape components, such as avalanches, wetlands, and areas affected by fire. This led to the analysis of scale dependencies and spatial phenomena, which is vital both for understanding ecosystem processes and for structuring the GIS. Current work focuses particularly on the construction of models of alpine tree line, especially in relation to topoclimatic factors and snow cover. Pulsford and Ferrier (1993) present a complementary paper which focuses on the implementation of a decentralized approach for the use of GIS by the National Parks and Wildlife Service of New South Wales, Australia. Case studies include the use of GIS in fire management and planning, surveys of fauna, wilderness assessment, archaeological studies, and the development of wildlife habitat links through areas proposed for logging.

The last of these topics indicates one of the major potentials of GIS: its usefulness as an objective methodology for developing management strategies in both protected areas and adjacent lands. The integrative value of GIS is especially valuable for resources and processes that cross the human-defined boundaries of protected areas: for instance, fires, animals, and people. Brown et al. (1993) give a concise account of the use of GIS for providing the information necessary to help

resolve conflicts between forestry and wildlife conservation interests in an area that includes logging leases and two National Parks in the Rocky Mountains of British Columbia, Canada. They show that overlay techniques in a GIS may be used to define population goals and habitat requirements for a viable caribou herd and develop trade-off scenarios for the modification of wood harvesting schedules. Thus, the GIS may be used as a neutral system for identifying potential conflicts through the development of scenarios for each of the stakeholders. This work has recently been extended to the economic evaluation of trade-offs between conserving forests for logging or as caribou habitat (Thompson et al., 1993). Such multiple accounts analysis is a major potential use of the data in GIS for economic decision-making which is just beginning to be realized.

Jordan (1993) also considers potential conflicts between different groups: in this case, the management personnel and villagers living in Sagarmatha National Park, Nepal. His study of deforestation risk suggests that some of the aspects of the theory of Himalayan degradation (Ives and Messerli, 1989) have been overstated. His simple model of deforestation risk, initially developed in England, was tested in the field, leading to an improvement in predictive value through field work and the application of local expert knowledge. This shows the importance of field testing of any results generated by deterministic GIS models, which are often presented – and even used for policy development – without field verification.

Simulation and Prediction

The use of GIS to suggest the future of managed and unmanaged mountain environments requires both the inclusion of the necessary range of variables at suitable spatial scales and a detailed understanding of their interactions. However, for many of the questions for which scientists and managers require answers, neither the data nor the theory are adequate. Thus, GIS may be conceived as systems that are valuable and not merely for recording, managing, and integrating data, but also for generating ideas and exploring relationships. Two exercises in this volume, Five and Six, are examples of using GIS to explore topographic relationships. In particular, the first explores relationships that might predict regional climatic patterns, and the second builds a model to predict the response of ecological patterns to topology.

The use of GIS for simulation and prediction is, at present, not very well-developed, especially for highly complex environments such as mountains, although many possibilities exist. An early use of simulation modelling for a small area (100 km²) was undertaken in the Swiss valley of Davos in the early 1980s (Binz and Wildi, 1986). In this work, a highly detailed GIS, with 50 m x 50 m grid cells, was used to evaluate the effects of various land-use scenarios – such as increased tourist development or the maintenance of an attractive cultural landscape – on avalanche hazards, as well as wildlife habitat, vegetation types, and other topics of interest for nature conservation.

One area of particularly great potential value is to use GIS to look back to the past and forward to the future. Such sequential analysis has been undertaken in the Middle Mountains of Nepal, showing a change from deforestation, in the late 1940s to the 1960s, to successful afforestation in the 1980s (Schreier et al., 1993). Over the same period, agriculture has been expanded onto more marginal sites. The combination of these two processes is leading to deficits in food sources for grazing animals. By incorporating a microclimatic model and expert knowledge into the GIS, the work has allowed the definition of the best uses for land in order to optimize the production of both trees and fodder. Such approaches might also be used for other diverse temporal processes, such as mass movements, the expansion of recreational impacts, and changes in the locations of glacier margins, permitting the simulation of ice dynamics.

Brzezicki et al. (1993) describe a major research project to simulate the impacts of climate change on the vegetation of Switzerland, a country that is nearly three-quarters mountainous. The project began with the integration of a DTM and environmental data sets as the basis for developing a model of potential natural vegetation on a 250 m x 250 m grid for the entire country. After quantitative and qualitative comparisons of the results of the model with “reality” (as represented on vegetation maps) showed a reasonable degree of reliability, experiments were conducted to assess likely effects of increasing temperature. This led to plausible results, with forest communities “moving” upwards as climatic conditions become more favourable.

Halpin (1993) presents a similar approach for the vegetation of Costa Rica, albeit at a coarser spatial scale, and also presents a comparative study of the impacts of climate change on the vegetation of a “hypothetical mountain” digitized into GIS of Alaska, California, and Costa Rica. He notes that previous approaches to assessing the effects of increasing tem-

perature on ecological zones in mountain used very simplistic rules, rather than GIS approaches based on detailed understanding of ecoclimatic relationships for asymmetric zones. The conclusions of past studies may therefore be not only simplistic but inaccurate; an important concern when they are applied to planning for protected areas, ecosystems, and species. The paper concludes that realistic plans for the management of mountain ecosystems require the integration of GIS analysis and evolving ecological theory, with validation through field work and remote sensing.

Discussion and Conclusions

Despite the recognized diversity of mountain environments and regions at all scales, it is possible to recognize a common set of issues and problems for the scientist and manager who wishes to apply GIS to management and research within these regions. In addition to the data issues discussed earlier, organisational and theoretical issues are also of great importance.

Organisational issues derive particularly from the marginal locations of most mountain regions. As they are generally not investment centres, funding for many activities, including the use of GIS, may be seriously restricted. Furthermore, political sensitivities may lead to restrictions on access to data. Accessibility in mountains is not only a political, but also a physical issue, relating to both the collection and the transfer of information in harsh, rugged environments. Given the common limitations of physical accessibility in mountain areas, decentralized approaches to the implementation of GIS, as described by Pulsford and Ferrier (1993), may be particularly applicable.

Perhaps a more important issue lies in whether or not current GIS technology provides an appropriate methodological framework in which to develop models about mountain environments. As mentioned earlier, the two-dimensional nature of virtually all current GIS places considerable constraints on the latitude within which one can develop models. This is perhaps typified by the significant errors that typically derive from the use of standard GIS functions, such as the calculation of area, slope, and aspect. This is not surprising when one considers the distortion necessary to transform “mountain space” into a planar projection. The “flattening” of the mountain environment creates further problems which, as noted above, lead to significant difficulties in creating land-cover maps from remotely-sensed imagery. For instance, critical environments with rare species, such as cliffs, may be effectively “invisible” in current GIS because their horizontal area is so minute. Equally, it is almost impossible to capture the micro-level diversity in the spatial organisation and arrangements of landscape features without the most detailed of surveys; although information about the optimum scale of analysis for specific purposes is increasingly becoming available (e.g., van Westen, 1993).

Turning to issues relating to data processing, we should note that all data processing algorithms have an implicit theory content. Data processing techniques – such as overlay, buffering, and image ratioing – contain our ideas about how mountain systems work. The important question is: how well does today’s GIS toolbox reflect our knowledge about the behaviour of mountain processes? Most GIS have an array of spatial processing functions which are combined in various ways to build a process model. The sceptic would argue that a GIS approach encourages us to seek an order and a simplicity which do not exist. Take, for example, the problem of migrating mammals and the spatial analysis of distance between habitats. In this case, a mountain area is best considered as an island with impassable barriers at high and, often, low elevations. Within this territory, animals will move depending upon the availability of food, cover, and other factors so that the distance function is a subtle, non-linear combination of the spatial arrangement of habitats. The algorithms available in most current GIS may be unable to treat such complex situations satisfactorily; at best it may be possible to make broad generalisations from satellite imagery and field work.

This leaves us with a fundamental dilemma: should we use present technology and data to develop a model which we know from the start will be inherently flawed, or should we seek an improved solution? We would argue that the challenge for mountain scientists is to take the latter course of action and to use existing GIS technology, explicitly recognizing its limitations, with the intention of improving both the contents of the toolbox and the methodological framework for their application. Nevertheless, there are undoubtedly some instances in which the resources required to obtain the hardware, software, and data necessary to create a GIS and then use it far outweigh the potential benefits for research or decision-making.

Finally, we should recognize another potentially serious problem: that the increased use of GIS may lead to scientists and

policy-makers spending more time in the office and computer laboratory rather than doing research and working in the field. This is a situation which could isolate both policy and research and lead to a negative view of a technology which, as shown by the papers in this volume, has much to offer for the understanding and management of mountain environments. Instead, we would promote the use of GIS as a means for integrating the understanding of natural and social scientists, the technological expertise of computer scientists, and the practical experience of managers and policy makers. As such, the value of GIS for mountain people and their environments can only grow.

References

- Alegre, P. (1982) *Una Aplicació del Programa M.A.P. a Catalunya*, Barcelona: Departament de Geografia, Universitat Autònoma de Barcelona.
- Ashdown, M. and Schaller, J. (1990) *Geographic Information Systems and their Application in MAB-projects, Ecosystem Research and Environmental Monitoring*, MAB-Mitteilungen 34, Bonn: German National Committee for the UNESCO Man and the Biosphere (MAB) Programme.
- Barry, R.G. (1992a) "Climate change in the mountains." In Stone, P.B. (Ed.) *The State of the World's Mountains*, London: Zed Books, pp. 359-380.
- Barry, R.G. (1992b) *Mountain Weather and Climate*, 2nd Edn, London: Routledge.
- Beissmann, H. (1989) *Plausibilitätsanalysen mit Hilfe eines EDV-gestützten Themakartographischen Informationssystems*, Berichte und Informationen 14, Vienna: Institut für Kartographie, Österreichische Akademie der Wissenschaften.
- Bird, A.C. and Taylor, J.C. (1990) "The role of GIS in monitoring landscape change in the National Parks of England and Wales." In Proceedings of 2nd National Conference of Canadian Institute of Surveying and Mapping, *GIS for the 1990s*, Ottawa, pp. 647-656.
- Binz, H.R. and Wildi, O., (1986) Szenarien, in Wildi, O. and Ewald, K. (Eds.) *Der Naturraum und dessen Nutzung im alpinen Tourismusgebiet von Davos*, Report 289, Birmensdorf: Swiss Federal Institute of Forestry Research, pp. 275-314.
- Brown, S.J., Schreier, H.E., Woods, G. and Hall, S. (1993) "A GIS analysis of forestry/ caribou conflicts in the trans-boundary region of Mount Revelstoke and Glacier National Parks, Canada." In Price, M.F., and Heywood, D.I. (eds.) *Mountain Environments and GIS*, London: Taylor and Francis, pp. 235-248.
- Brzeziecki, B. Kienast, F. and Wildi, O. (1993) "Potential impacts of a changing climate on the vegetation cover of Switzerland: a simulation experiment using GIS technology." In Price, M.F., and Heywood, D.I. (eds.) *Mountain Environments and GIS*, London: Taylor and Francis, pp. 263-280.
- Bucek, A. and Lacina, J. (1981) "The use of biogeographical differentiation in landscape protection and design. " *Sbornite* (CSGS, Praha), 86(1): 4-50.
- Burkmar, R., et al. (1992) "Integrating GIS and knowledge-based systems for landscape management." *Proceedings, Conference of the Association of Geographic Information, Birmingham*, pp. 1-5.
- Butler, D.R., Walsh, S.J. and Malanson, G.P. (1990) "GIS applications to the indirect effects of forest fires in mountainous terrain." In *Fire and Environment: Ecological and Cultural Perspectives*, General Technical Report SE-69, Washington DC: US Department of Agriculture, Forest Service, pp. 202-211.
- Carver, S., Heywood, D.I., and Cornelius, S. (1993) "Evaluating field-based GIS for environmental characterization." *Proceedings of the 2nd International Conference on Integrating Geographic Information Systems and Environmental Modelling*, National Centre for Geographic Information and Analysis, Breckenridge, Colorado.
- Claudin, J., Sourp, E. and Puydarrieux, P. (1993) "Constitution d'un système d'information géographique pour la gestion d'un espace naturel: l'expérience du parc national du Mercantour." in *Environmental Information Systems*, Le Bourget-du-Lac: International Centre for Alpine Environments, pp. 31-42.

- Culbertson, K. Hershberger, B., Jackson, S. Mullen S. and Olson H. (1993) "Geographic information systems as a tool for regional planning in mountain regions: case studies from Canada, Brazil, Japan, and the USA." In Price, M.F., and Heywood, D.I. (eds.) *Mountain Environments and GIS*, London: Taylor and Francis, pp. 99-118.
- d'Oleire-Oltmanns, W. (1987) MAB-Projekt 6 Habitatbewertung und potentielle Verbreitung von Tierarten unter touristischer Einfluss, *Verhandlungen der Gesellschaft für Ökologie* (Graz 1985), 15: 48-56.
- d'Oleire-Oltmanns, W. (1990) "The interaction of patchiness, land cover type and animal distribution: an evolution in time and space." *Proceedings, Resource Technology 90*, Washington D.C., pp. 369-375.
- d'Oleire-Oltmanns, W. and Franz, H.P. (1991) "Das zoologische Informationssystem (ZOO LIS) der Nationalparkverwaltung Berchtesgaden." *Verhandlungen der Gesellschaft für Ökologie* (Freising-Weihenstephan 1990), 20: 685-692.
- d'Oleire-Oltmanns, W., Franz, H.P. and Schuster, A. (1991) "Die Anwendung des Ökosystemforschung für die Analyse der räumlichen Habitatverteilung von Tierarten." *Verhandlungen der Gesellschaft für Ökologie* (Osnabrück 1989), 19(3): 619-627.
- Downey, I.D., Heywood, D.I., Kless, P., Pauknerova, E. and Petch, J.R. (1990) "GIS for landscape management: Zdarske Vrchy, Czechoslovakia." In Proceedings of 2nd National Conference of Canadian Institute of Surveying and Mapping, *GIS for the 1990s*, Ottawa, pp. 1528-1538.
- Downey, I.D., Heywood, D.I. and Petch, J.R. (1991) "Landscape ecology as an operational framework for environmental GIS." In Jackson, M.C. (Ed.) *Systems Thinking in Europe*, New York: Plenum, pp. 151-158.
- Forman, R.T.T. and Gordon, M. (1986) *Landscape Ecology*, New York: Wiley.
- Gerrard, A.J. (1990) *Mountain Environments*, London: Belhaven.
- Grötzbach, E.F. (1988) "High mountains as human habitat." In Allen, N.J.R., Knapp, G.W. and Stadel, C. (Eds.) *Human Impact on Mountains*, Totowa, New Jersey: Rowman and Littlefield, pp. 24-35.
- Haefner, H., Gresch, P., Hugentobler, F. and Marti, S. (1991) "Landschaftswandel und Landschaftsschutz: Das Beispiel des Malcantone, Kanton Tessin." *Regio Basiliensis*, 32(2): 125-135.
- Haefner, H., and Hugentobler, F. (1985) *Assessment and Monitoring of Abandoned Agricultural Land in the Swiss Alps—Methods and Examples*, Remote Sensing Series 9, Zürich: Department of Geography, University of Zürich.
- Haefner, H., and Hugentobler, F. (1988) "Monitoring of natural resources for environmental and regional planning in the southern Swiss Alps." *Proceedings, 8th EARSeL Symposium*, Capri, pp. 368-377.
- Halpin, P. (1993) "A GIS analysis of the potential impacts of climate change on mountain ecosystems and protected areas." In Price, M.F., and Heywood, D.I. (eds.) *Mountain Environments and GIS*, London: Taylor and Francis, pp. 281-301.
- Haslett, J.R. (1990) "Geographic information systems: a new approach to habitat definition and the study of distributions." *Trends in Ecology and Evolution*, 5: 214-218.
- Haslett, J.R. (1993) "Complicated habitat mosaics and insect community structure on mountains." *Proceedings of the Royal Society B*, in press.
- Hewitt, K. (1992) "Mountain hazards." *GeoJournal*, 27: 47-60.
- Ives, J.D. (1992) Preface, in Stone, P.B. (Ed.) *The State of the World's Mountains*, London: Zed Books, pp. xiii-xvi.
- Ives, J.D. and Messerli, B. (1989) *The Himalayan Dilemma: Reconciling Development and Conservation*, London: Routledge.
- Jordan, G. (1993) "GIS modelling and model validation of deforestation risks in Sagarmatha National Park, Nepal." In Price, M.F., and Heywood, D.I. (eds.) *Mountain Environments and GIS*, London: Taylor and Francis, pp. 249-260.

- Koshkariov, A., Krasovskaia, T.M. and Tikunov, V.S. (1993) "Towards resolving the problems of regional development in the mountains of the Commonwealth of Independent States using Geographic Information Systems." In Price, M.F., and Heywood, D.I. (eds.) *Mountain Environments and GIS*, London: Taylor and Francis, pp. 77-98.
- McKendry, J.E. and Eastman, J.R. (1992) "Applications of GIS in forestry: A review." In McKendry, J.E., Eastman, J.R., St. Martin, K. and Fulk, M. *Applications in Forestry*, UNITAR Explorations in Geographic Information Systems Technology, Vol. 2, Clark University, Worcester, Mass.
- Michal, I. (1992) *Ecologica Stability*, Brno: Veronica.
- Miller, D.R., Morrice, J.G., Horne, P.L. and Aspinall, R.J. (1993) "Use of GIS for analysis of scenery in the Cairngorm mountains of Scotland." In Price, M.F., and Heywood, D.I. (eds.) *Mountain Environments and GIS*, London: Taylor and Francis, pp. 119-132.
- Parker, C.R., Langdon, K., Carter, J.R., Nodvin, S.C. and Barrett, H. (1990) "Natural resources management and research in Great Smoky Mountains National Park." *Proceedings, Resource Technology '90, 2nd International Symposium on Advanced Technology in Natural Resource Management*, Falls Church: American Society for Photogrammetry and Remote Sensing, pp. 254-264.
- Parrachini, M.L. and Folving, S. (1993) "Land use classification and regional planning in Val Malenco (Italian Alps): a study on the integration of remotely-sensed data and digital terrain models for thematic mapping." In Price, M.F., and Heywood, D.I. (eds.) *Mountain Environments and GIS*, London: Taylor and Francis, pp. 59-76.
- Pauknerova, E., Brokes, P., Heywood, I., and Petch, J. (1992) "GIS and remote sensing team-up in Bohemia-Moravian Highlands study." *GIS Europe*, 1(4): 18-21.
- Petch, J.R., and Kolejka, J. (1993) "The tradition of landscape ecology in Czechoslovakia." In Haines-Young, R. Green, D.R. and Cousins, S.H. (eds.) *Landscape Ecology and GIS*, London: Taylor and Francis, pp. 39-56.
- Pulsford, I., and Ferrier, S. (1993) "The application of GIS by the National Parks and Wildlife Service of New South Wales, Australia to conservation in mountain environments." In Price, M.F., and Heywood, D.I. (eds.) *Mountain Environments and GIS*, London: Taylor and Francis, pp. 217-234.
- Quarrie, J., Ed. (1992) *Earth Summit '92*, London: Regency Press Corporation.
- Rodcay, G. (1991) "GPS/GIS project benefits biodiversity." *GIS World*, 9: 52-53.
- Schaller, J. (1993) "GIS and ecosystem models as tools for watershed management and ecological balancing in high mountain areas: the example of ecosystem research in the Berchtesgaden area, Germany." In Price, M.F., and Heywood, D.I. (eds.) *Mountain Environments and GIS*, London: Taylor and Francis, pp. 43-58.
- Schmidt, M. and Schreier, H. (1991) "Quantitative GIS analysis of the forest resources in a mountain watershed in Nepal." *GIS-91, Applications in a Changing World*, Vancouver: Forestry Canada, pp. 227-231.
- Schreier, H., Shah, P.B. and Kennedy, G. (1990) "Evaluating mountain watersheds in Nepal using micro-GIS." *Mountain Research and Development*, 10: 151-159.
- Schreier, H., Brown, S., Schmidt, M., Shah, P.B., Shrestha, B., Nakarmi, G., and Wymann, S. (1994) "Gaining forests but losing ground: A GIS evaluation in a Himalayan watershed." *Environmental Management*, 18(1): 139-150.
- Schuster, A. (1990) "Ornithologische Forschung unter Anwendung eines Geographischen Informationssystems." *Salzburger Geographische Materialien*, 15: 115-123.
- Schweinfurth, U. (1992) "Mapping mountains: vegetation in the Himalaya." *GeoJournal*, 27: 73-83.
- Simonett, O. (1993) *Geographic Information Systems for Environment and Development*, Zürich: Geographisches Institut, Universität Zürich-Irchel.

- Steiner, D. and Zamani, F. (1982) *Datenbank MAB-Grindelwald*, Fachbeitrag zum schweizerischen MAB Programm 21, Bern: Bundesamt für Umweltschutz.
- Stocks, A.M. and Heywood, D.I. (1993) "Terrain modelling for mountains." In Price, M.F., and Heywood, D.I. (eds.) *Mountain Environments and GIS*, London: Taylor and Francis, pp. 25-40.
- Thompson, W.A., Van Kooten, G.C., Vertinsky, I., Brown, S. and Schreier, H. (1993) "A preliminary economic model for evaluation of forest management in a geo-referenced framework." *Proceedings, 7th International Symposium on Geographic Information Systems in Forestry, Environment and Natural Resource Management*, Vancouver: Forestry Canada and Polaris Learning, pp. 529-539.
- Trafas, K., Ed. (1985) *Atlas of Tatra National Park*, Zakopane and Cracow: Tatra National Park and Cracow Section of the Polish Association of Earth Sciences.
- van Wagtenonk, J.W. (1990) "GIS applications in fire management and research." *Fire and environment: ecological and cultural perspectives*, General Technical Report SE-69, Washington DC: US Department of Agriculture, Forest Service, pp. 212-214.
- van Westen, C.J. (1993) "GIS in landslide hazard zonation: a review, with examples from the Colombian Andes." In Price, M.F., and Heywood, D.I. (eds.) *Mountain Environments and GIS*, London: Taylor and Francis, pp. 135-166.
- Vasconcelos, M.J., Pereira, J.M.C. and Zeigler, B.P. (1993) "Simulation of fire growth in mountain environments." In Price, M.F., and Heywood, D.I. (eds.) *Mountain Environments and GIS*, London: Taylor and Francis, pp. 167-186.
- Wadge, G. (1988) "The potential of GIS modeling of gravity flows and slope instabilities." *International Journal of Geographic Information Systems*, 2: 143-152.
- Walsh, S.J., Butler, D.R., Brown, D.G. and Bian, L. (1993) "Form and pattern in the alpine environment: an integrated approach to spatial analysis and modelling in Glacier National Park USA." In Price, M.F., and Heywood, D.I. (eds.) *Mountain Environments and GIS*, London: Taylor and Francis, pp. 189-216.
- Walsh, S.J., Cooper, J.W., Von Essen, I.E., Gallager, M.R. (1990) "Image enhancement of Landsat Thematic Mapper data and GIS data integration for evaluation of resource characteristics." *Photogrammetric Engineering and Remote Sensing*, 56: 1135-1141.
- Weibel, R., and Heller, M. (1991) "Digital terrain modelling." In Maguire, D.J., Goodchild, M.F. and Rhind, D.W. (Eds.) *Geographic Information Systems*, London: Longman, pp. 269-297.
- Willison, J.H.M. et al. (1992) *Science and Management of Protected Areas*, Amsterdam: Elsevier Science Publishers.
- Yozon, P., Jones, R. and Fox, J. (1991) "GIS for assessing habitat and estimating population of red pandas, Langtang National Park." In Schreier, H., Brown, S. and O'Riley, P. (Eds.) *Proceedings, 5th International Symposium on Geographic Information Systems in Forestry, Environment and Natural Resources*, Vancouver: Forestry Canada and Reid Collins Ltd., pp. 233-244.

Exercise 1: Digital Terrain Modeling: An Overview



Introduction

Mountainous terrain greatly influences the climate and vegetation of locations across the earth's surface. It also influences a variety of human activities, especially those occupations that are directly related to the land. Planners, resource managers, and environmentalists working in the fields of hydrology, soils, geomorphology, and ecological research often encounter the need for topographic data. Using digital terrain models (DTMs), the analyst incorporates topographic relationships in tasks such as surface characterization, site visibility analysis, wildlife habitat mapping, hazards prediction, satellite imagery classification, and modeling human and environment interactions. A digital terrain model can be used both as a means of modeling the three-dimensional character of an area and as a base for deriving other characteristics of mountain environments (Heywood et al., this volume). The development of a DTM, its data structure, scale, resolution, and quality thus become special considerations of a GIS project. In this exercise, we explore GIS techniques for modeling characteristics of the earth's surface.

We begin here by examining a traditional contour map and considering its form and limitations in computer applications. We then examine the differing solutions to the problem of representing 3-dimensional surfaces and elevations in GIS. Vector Triangulated Irregular Networks (TINs) and raster Digital Elevation Models (DEMs) are each forms that possess inherent strengths, weaknesses and application possibilities.

For vector systems, the Triangulated Irregular Network is a model which creates a topographic surface by connecting measured points of elevation with lines of slope and planes of aspect. We will examine a TIN model here and consider levels of detail and distortion in such models.

We then move on to raster DEMs and the steps in their creation. Though several methods exist for the creation of a raster elevation surface, in this exercise we provide one of the most common examples by creating a DEM through the interpolation of contours. These contours are created by digitizing topographic data and transforming the resultant vector file into a raster format. Raster interpolation is a procedure for estimating values at locations which are not part of the original data set. It fills in the gaps among data to make a full elevation surface or grid. In the process, issues of interpolation method and resolution inevitably arise.

The model type we chose and the level of detail of that model will affect data quality and any additional information derived from it through geographical analysis. These issues are explored here for a raster data model in particular. Errors and accuracy in interpolation depend upon the selection of an appropriate cell size for the topographic data set. A quantitative method for determining resolution is provided here. An application case study is then examined to show the implications of cell resolution on data derived from a DEM. As will be shown, varying cell resolutions will result in widely varying data output for decision making.

Data for the exercises cover a small portion of the Salzkammergut region of Austria and of Middle Himalayas in Nepal. The data set includes the following:

AUSDEM	a digital elevation model, the Austrian Alps
AUSTIN	a triangulated irregular network of the same region
AUSTIN2	a more detailed TIN covering the same region as AUSTIN
HIMAL	a vector file of contours (250 feet interval), Himalayas
RASHIMAL	a raster version of HIMAL

Topographic Data Models

A digital terrain model is a digital representation of relief, terrain, or ground surface. It consists of three dimensional data made up of length, width, and height measurements. These can be projected upon a two dimensional referenced surface like a computer screen, through a number of models. Often, the original relief is derived from a measurement of heights on the earth's surface taken from irregularly spaced locations known as control points. Such data is acquired through one or a combination of the following methods:

1. Direct surveying of the terrain by tachometry and global positioning systems.
2. Photogrammetric data acquisition from stereo pairs of aerial photographs.
3. The digitizing and/or scanning of existing maps.
4. Automatic generation of contours from satellite data (e.g., SPOT).

Many photogrammetric techniques make use of grids to provide a basis for generating relief data. The grid method represents data in a predictable order of cells arranged by columns and rows. It is the most widespread today partly because of its compatibility and ease of use in computer analysis.

- a) First, open IDRISI and under the File menu, select IDRISI Explorer. Then select it's Projects tab. This tab allows you to set the project environment of your file folders. Make sure that the Editor pane is open at the bottom of the form. If you right-click anywhere in the Projects form, you can select to show the Editor. The Editor pane shows the working and resource folders. Choose to either create a new project or edit the existing project to include the folder holding the data for this exercise, (e.g., C:\UNITAR\Mountain\Data\Exer1). Please note that this folder must be listed as either the main Working folder or one of the Resource folders. Also open the User Preferences dialog from the File menu and click on the Revert to Defaults button, then click OK.

To observe the difference between older cartographic methods of representing elevation versus newer digital methods, run DISPLAY Launcher from the Display menu. Select to display a vector file. Bring up the list of file-names by clicking on the browse button (the button just to the right of the file input box) or by double-clicking in the input box. Select the file HIMAL and click OK. Retain the standard default symbol file and click OK. This image represents a portion of a digitized 1:50,000 topographic map for the Dhading District of Nepal¹. This is a direct digital representation of the paper map.

Traditional cartographic methods utilize the contour, as displayed on the screen. This consists of a line which joins points on the earth's surface that are of equal height. The spacing of different contour lines reflects the gradient of slope on the represented surface. Closely-spaced contour lines depict steep slopes while widely spaced contours represent gently sloping surfaces. Plane surfaces, such as the central portions of the image on the screen, rarely have contour representations. They could be flat areas occurring at valley bottoms or at the tops of mountains. Points represent elevation peaks.

This image, like other forms of cartographic representation, conveys information visually, but allows no further digital analysis. For this reason, GIS typically represents elevation by means other than traditional cartographic methods. These other data forms support the analytical and modeling capabilities of computers. Computers enable us, for example, to do analytical hill shading. This is a means of enhancing a three-dimensional view by simulating sun illumination at an angle across the surface.

In this exercise, DTM refers to models of either vector or raster. The different data structures affect the types of analyses possible and the kinds of information derived from these models. Because vector and raster systems use different underlying data structures, they employ different representations of altitude, slope, and aspect. In the case of vector systems, the Triangulated Irregular Network (TIN) is used. Raster systems employ a grid-based Digital Elevation Model (DEM).

1. The digital data was contributed by the MENRIS staff of ICIMOD in Kathmandu, Nepal.

- b) Close the display of the file HIMAL.

Vector Triangulated Irregular Networks (TIN)

The TIN data model is a conceptually simple interpolation in which irregularly spaced points of measured elevation are connected by lines and planes. By connecting every elevation point with its neighbors, a network of triangles is formed. These are used to create a smooth and continuous surface which approximates the terrain from which the original data points were gathered. This method of representation is particularly good for alluvial landscapes broken at sharp angles and along ridges (Mark, 1975).

TIN models are based on applying Delaunay triangulation to point elevation data. This method forms connected but non-overlapping triangles such that the perimeter of each triangle contains no other point in its interior. The vertices represent elevations, lines hold slope information and polygons possess an associated aspect value (See Figure 1).

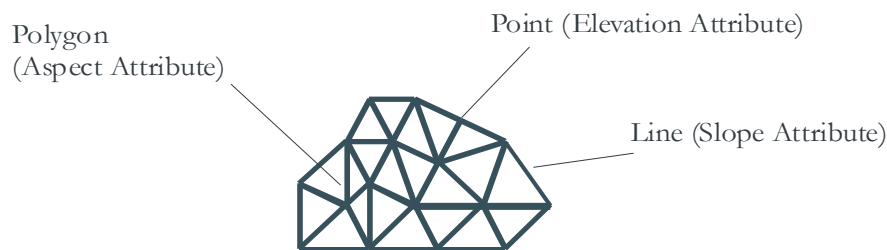


Figure 1: Topographic Information in a TIN

- c) Run DISPLAY Launcher, but this time do not access it from the Display menu. Instead access it from its toolbar icon (the one of the world map, second from the left). Select to display a vector file. Type in the filename AUSTIN and click OK.² The TIN data structure consists of a mesh of lines composed of triangles in which each vertex (point feature) is a site with an assigned height ("z") measurement. The sides of the triangle (line features) have specific slope values. Each triangle (polygon) has an associated aspect value. All of this information is stored in a table that contains values that link the table to the TIN features.

Run DISPLAY Launcher once more and select to display a vector file named AUSTIN2. Click OK to accept the defaults (the default symbol file is Quantitative). Note the different sizes of triangles used to represent the surface details in the map. This is a TIN model for the same area as AUSTIN but in this case, the number of sampled elevation points has been increased to capture the surface in more detail. (Note that these are representations of the TIN model; we do not have elevation, slope, and aspect information registered with these features.)

Vector contours are discontinuous; no information is provided for the spaces between contours. They do not provide enough information for describing terrain features. The TIN technique is thus an improvement over simple contours stored in vector format. Currently, it is used by vector software designers to support three dimensional visualization, to derive slope and aspect information from elevation data, and to do visibility analyses. Further, the TIN structure builds surfaces which allow the use of topographic information alongside other variables in geographic analysis and decision-making (e.g., locating a waste facility). Finally, this method does not store intermediary values of heights across the surface and so requires less computer memory than other digital terrain models.

2. Dr. Josef Strobl, Salzburg University, Salzburg, Austria contributed the entire Austrian data set which includes the TIN data as developed in Arc/Info, an interpolated raster DEM, and contour lines. All Austrian images are of the same geographic area.

The main advantage of using the triangular network technique for mapping relief, lies in the consideration of accuracy in surface representation and speed of computation (McCullagh, 1988). The TIN retains surface detail which is favorable for applications concerned with the linear details or surface shapes occurring on the terrain. To capture the same level of detail in raster may be far more data intensive than is desirable. The TIN model also can be directly implemented from a triangulation survey using the original points on the relief surface.

The TIN, like any interpolation procedure, confronts the problem of orienting data features (point, line, triangle or pixels) to best describe the terrain. The primary problem of the TIN interpolation method is in the allocation of the triangles to best describe the terrain. This will depend on both the distribution and the density of sampled elevation points. If the points are collected at regularly and equally-spaced locations, the resulting TIN generally does not produce satisfactory results. If the distribution of points does not reflect locations of inflection (ridges, peaks, and pits), then the ability of the TIN to represent the surface decreases.

Similarly, the density of information recorded for heights must vary according to the rate of change occurring on the earth's surface. Some terrain locations in a region require more elevation points for accurate representation than others. Point selection and sampling therefore become important in TIN creation. Methods have been developed which search for and sample important surface-specific points on the represented surface like peaks or pits (Chen and Guevara, 1987). Considerations like these reflect an important principle: a greater number of sample points may not lead to a more accurate representation of the surface if they are poorly located. There is therefore not always a direct relationship between detail and accuracy.

In summary, the TIN provides a digital topographic surface for analysis which is easy to interpret and not too data intensive. This method of interpolation does have certain limits in analytic power however, and as explained above, lends itself to certain kinds of distortions.

As is discussed in the next section, raster interpolation algorithms are similarly applied to create surfaces from known data points or contours. Like TINs, these are calculated across features that change in shape at variable rates. The design of interpolation methods in either case tries to minimize two common kinds of distortions: those having a tendency to occur in areas of either rapid or gradual change in relief.

Raster Digital Elevation Models (DEM)

There are several methods for producing a raster DEM. One method derives a DEM from a TIN data model. Linear equations describing each triangle are used to calculate the elevation value occurring at each pixel in the corresponding raster grid. The following is an example of a raster model developed from a TIN.

- d) Use DISPLAY Launcher to view the image file AUSDEM. Click OK to accept the defaults for the remaining options. While the image is still displayed on the screen, select the Cursor Inquiry Mode icon (the one with a question mark and arrow) to activate the cursor. Move the cursor to a location on the image. Note that a cross shaped marker appears and that its column, row, X, and Y position are shown at the bottom of the screen. To check the data value of the pixel at the cursor's location, press the left mouse button once. A "z" value will be displayed next to the cursor. The "z" value in this case corresponds to the elevation value. Move the cursor to another location and press the left mouse button again to check this location's elevation value. When done, exit Cursor Inquiry Mode by pressing the toolbar icon so that it is no longer highlighted.

Notice that elevations are continuously changing. This is an image interpolated from the TIN model represented in file AUSTIN2.

- e) Next, zoom in close enough to see the cell-like structure of the raster grid.

Check the values again to see how they are changing.

- f) Next, we will examine the display of the study area in an orthographic or three-dimensional perspective. Run ORTHO from the Display menu. Specify the surface image as AUSDEM2.³ Click on the Drape image box to

uncheck it. Click OK to accept the defaults for the remaining options. When done viewing the image, close it.

In a raster system, space is represented as a rectangular grid with each grid cell (pixel) containing a data value. Every cell in this rectangular space holds information. This allows for a continuous representation of an area as a surface rather than as a set of boundaries or point features. Mathematical procedures are used to interpolate elevation values between pairs of contours or measured point heights. The result is a continuous *surface* of elevation.

The column and row structure of the raster grid is advantageous since it makes locations predictable and is inherently compatible with the computer's analytical processes. Because the computer knows the location of pixels relative to one another, it can derive new information for each pixel based on its relationship to neighboring pixels. For example, by interpreting the relationship between neighboring elevation values as slope, the computer can quickly calculate a slope surface from the DEM (See Figure 2). Aspect may similarly be derived through the application of a different spatial algorithm to the DEM. The results of the calculation may then be stored as an independent data layer in which each pixel represents a value of slope. This allows easy integration of topographic data with other layers of information. Employing a digital elevation model in a raster system therefore opens up greater possibilities for modeling relationships and processes occurring in the landscape, especially those affected by topography.

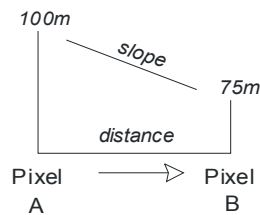


Figure 2: Slope Calculation from a DEM

There are constraints, however, on the accuracy of the final interpolated DEM. This is true of any sampled representation of a real surface. The first concern is cell size (or resolution). A pixel represents a single height for a square *area* rather than a *point*. As such, the larger the size of the pixel, the more generalized is the estimate of height for that space. With rapidly changing altitudes, too large a resolution poses significant problems of accuracy, depending upon the purpose of the DEM.

A second concern is the method of interpolation. As was mentioned in the discussion of TIN models, the quality of altitude sampling and the type of interpolation method affects the kinds of distortions that are likely to occur. Using the same input data set, different interpolation functions may result in markedly differing terrain models. There are a variety of methods available. These fall into many general sets; global and local, exact and approximate, stochastic and deterministic are all possible characterizations of interpolation methods. For a more thorough discussion of the differences, see Lam (1983) and McCullagh (1988).

Global "approximate" methods fit an interpolation function to the entire raster data set such that all the data are handled in exactly the same manner by a polynomial function that creates the surface. These methods, like INTERCON in IDRISI, use an exponential (linear, quadratic, or cubic) interpolation function to estimate values between contours or points. Such a function has a tendency to preserve general shapes of steep terrain and gradually changing surfaces. It is a method which provides a good solution to the interpolation problem for a variety of general applications. Importantly, it does not make high demands on computing time and is available in a number of GIS packages.

Unlike TINs, global polynomial interpolation methods create smooth surfaces which may or may not reflect the high-

3. AUSDEM2 was created from AUSDEM by reducing the number of columns and rows by half (using the CONTRACT module). The orthographic display improves when the data set is thinned.

angle breaks in terrain sometimes found in earth surfaces. Another problem associated with this method is a tendency towards coarseness in the output surface. While general shape is maintained, the surface may have a rough or bumpy quality not seen in the topography being simulated. Filtering operations are used to smooth surface roughness occurring from this interpolation.

More complex (sometimes labeled "exact") global methods might also be employed. This group includes Kriging methods. Kriging methods are based on the assumption that the variance between sample points is not uniform and changes depending upon the distance between the points. This variation is represented in a *variogram* which describes the changes in difference in terms of distance. This variogram is derived from the original sample points and is used to create differential weights for the interpolation. It creates a surface with a minimum amount of variance based upon the input data. The method is analytically complex and may be unwieldy for many computers (Lam, 1983; Burrough, 1986). Only simple Kriging applications are available for GIS at present. These do not as yet make full use of the mathematical potential of the application. Several geostatistical packages are available however, which contain full Kriging capabilities. These may be highly compatible with GIS.

Other interpolation methods include "local" methods which apply different algorithms to different parts of a region. There are also "areal" interpolation methods which interpolate whole regions or polygons rather than individual points.

All interpolation functions vary according to their ability to relate the original sampled data set to the geometry of the surface terrain. They can all potentially create either false features or overly generalized surfaces. In the final analysis, the choice of method is very much dependent upon the user's preference, needs, time, costs, and available data.

In this exercise, we will focus on the development of a DEM from contours. This is a likely method of DEM construction since contour data are broadly available, and the interpolation methods are among the most common and useful in GIS. The contours can simply be digitized from topographic relief maps and transferred into raster format. Though digitized and rasterized with little difficulty, contours do present problems during interpolation. Some of these problems are summarized below:

- 1) The scale and contour interval of the original map from which the contours are digitized will set the limits on the accuracy of the resulting raster DEM. No subsequent increase in pixel resolution will improve the accuracy of a coarse or small scale map based image. This is not limited to raster applications, and should be kept in mind in any GIS application using data incorporated from hard copy maps.
- 2) Considerable computer time is needed to interpolate a detailed regular grid. This is because the grid interpolation process has a speed proportional to the number of cells in the output image.
- 3) Contours do not incorporate all of the information of a relief surface. Also, a raster pixel that describes features as square areas may not best represent features like break-lines, ridge and valley lines, and mountain peaks.
- 4) The global method used in interpolation lacks flexibility in responding to variable data densities that may occur in different portions of a relief map. There are problems of applying an algorithm uniformly to an area that contains two types of surfaces: one with rapidly changing relief and another of gradual change. Some algorithms are designed to function best with local variation and others with broad surface shapes. To minimize irregularities appearing in interpolation results, additional information may be required at specific places like hill tops and valley bottoms.

Again, the best application of any interpolation method depends on the use of the most appropriate sampling method (points, lines, networks) for that interpolation technique. There is no single best method for sampling or interpolation for either vector or raster terrain representation.

DEM Spatial Resolution Issues

The quality of available data and the suitability of that data for GIS in mountain environments determines the limits of possible applications. It is necessary to understand what data features will affect analysis, and to choose data representa-

tions most appropriate for that analysis. In particular, it is important to understand the effect of resolution in order to maximize the potential of available data.

As stated earlier, raster systems store data and display features in a regularly-spaced grid of cells or pixels. They are therefore not as well suited for the display of edges, boundaries, and points of features as vector systems are. In order to display linear features such as cliff lines or contours, a very fine cell resolution is necessary.

A few examples will clarify the distinction between raster and vector representations of features.

- g) Display the vector file HIMAL with the standard default palette. Note the nature and speed of the display on the screen. Next, display the image file RASHIMAL with the default Quantitative palette. Notice once more the nature of the display. The second image has a spatial resolution of 60 meters. This is a raster representation of the contour map (HIMAL). Click on the Add Layer button in Composer to overlay the vector file HIMAL on the image with the Qualitative symbol file. This file reflects the original lines of contour. Finally, use the Zoom Window tool to zoom into the upper left quarter of the image.

It is apparent that the low resolution of the underlying raster structure has distorted the display of the original data. The wide raster cells follow the vector lines of contour in a rough or boxy fashion. To better achieve the appearance of these vector contours in a raster system, we would need to increase the resolution by decreasing the pixel size. This would consequently increase the number of columns and rows to potentially unmanageable proportions.

The purpose of rasterizing contour lines is not generally for the display of the contour lines on their own however, but for their interpolation into a full elevation surface. Spatial resolution will still significantly affect how generalized the resulting surface will be.

To proceed with further analyses using the contour data in IDRISI, we need to convert HIMAL from a vector to a raster format. To rasterize vector files in IDRISI, we begin by creating "blank" raster grids into which we will deposit the vector data. These must be created with the row and column structure and the spatial resolution with which we want to work.

- h) To perform this conversion, you will run the module RASTERVECTOR from the Reformat menu five times using the line to raster option. You will create five separate images with the following output image names: CONT90, CONT60, CONT30, CONT15, and CONT7. The numbers represent the pixel resolution of each image. Using RASTERVECTOR, enter HIMAL as the vector line file. Enter the name of the output image you are creating (such as CONT90) as the image file to be updated. Click OK. Click Yes when prompted with the message "Bring up INITIAL to create this image?" In the INITIAL dialogue box, choose the option to define spatial parameters individually. Choose Integer as the output data type, and leave the initial value as 0. Click on the Output Reference Information button. Enter the number of columns and rows for your new image from the table below. Notice that IDRISI supplies the X and Y coordinates, reference system, reference units, and unit distance. Click OK, and then click OK again on the main INITIAL dialogue box to perform the rasterizing operation.

	CONT90	CONT60	CONT30	CONT15	CONT7
Columns :	90	135	271	581	1159
Rows :	82	122	240	488	1046

1. *How are these columns and rows determined?*

- i) Examine the image file CONT90. Use Add Layer in Composer to overlay HIMAL on this image with the Qualitative symbol file. Examine closely the spacing of the contours relative to the pixels. Notice that some pixels have two contour lines passing through them.

2. *If the contour interval is 250 feet, what does this suggest about the change in elevation relative to the pixel resolution?*

This is a form of distortion. When there is more than one contour crossing the area of a pixel, the line rasterization process arbitrarily assigns to the pixel the value of the contour that last crosses it, according to the order the contours are written in the vector file. At a very coarse resolution, this may not be a problem if the model is used only for general visualization. The interpolation of these contours will not likely produce severe irregularities if there is little or no blank pixel space between data cells with contours. It will produce an exaggerated smooth surface however. While safe for visualization, any model that summarizes the surface in this way is problematic when it is the basis for making decisions about resources and people. If the model were applied to the development of slope stability programs for example, any over or underestimation of slope derived from this model could cost money or even lives.

- j) Display CONT60 with the default Quantitative palette. Window in on the upper left quadrant of the image using the Zoom Window tool and the mouse. Overlay HIMAL on the image using the Add Layer option of Composer with the Qualitative symbol file. There is some improvement in the allocation of contour values relative to the resolution. Notice, however, that some raster contours have breaks and discontinue. This illustrates again the problem of overlapping; the resulting model will exaggerate the smoothness of the surface in areas of overlap
- k) Next, display CONT30 with the default palette. Window into the upper left corner of the image. Notice that more terrain detail appears. Find areas where contour lines overlap. This suggests that the rate of change in altitude in some locations is still greater than 250 feet per 30 meters.

While images at higher resolution better reflect the rate of topographic change, overlapping contours give rise to other potential interpolation problems. Because the amount of blank pixel space between contours has increased, there is more room for errors to become visible on the interpolation. With more pixels to interpolate, any errors due to overlapping lines become more severe.

- l) Display CONT15 with the default palette. Most problems disappear at this resolution. Zoom in on the upper left corner with the Zoom Window tool. The problem of touching lines still exists. One limitation of the raster grid structure is accommodating for rapid changes (steep slopes) in the representation of relief.
- m) Finally, examine the image CONT7. At this resolution, the contour lines lose their apparent raster grid-like character and the lines give the illusion of looking like vector features. Problems of overlap have disappeared. This resolution does not, however, contribute any new information for the interpolation process and will increase processing time for any application by adding a mass of redundant cells. Close all images before continuing to the next section.

Interpolation of Contours

The next step in the creation of the DEM is the interpolation of the rasterized contour lines. In IDRISI, this procedure is carried out through the module INTERCON. To initiate the algorithm, INTERCON requires elevation data for the four corners of the image.

- n) To record these elevations, display CONT90 with the default palette. Use the Zoom Window tool to window in on the upper left corner of the image. Once you have zoomed in, use the arrow keys to zoom closely enough in order to clearly view the contours closest to the first column and row position. Next, select Cursor Inquiry Mode and move the cursor on the contour pixel located at column and row 0,0. Press the left mouse button once and write down the elevation value ("z"). Window back out to the whole image by pressing the Home key on your keyboard. Repeat the procedure to record the elevation at the top right, bottom left, and bottom right. Estimate the elevation where no contour value is found at a corner pixel.
- o) Next, run the INTERCON module from the GIS Analysis/Surface Analysis menu. Specify CONT90 as the input image to interpolate and call the output image DEM90. Provide the program with the elevation estimates recorded for the four corners of the image and click OK. We will run INTERCON again for smaller resolutions. Before doing so, examine the corners of CONT60 and CONT30 to determine if the change in resolution affects your estimate of corner elevation values. Then, follow the same procedure as above to create DEM60 from

CONT60 by running INTERCON. Be aware that the length of time it takes for the program to interpolate increases exponentially with the number of cells in the file. For this reason, DEM30 and DEM15 have already been created in advance for you.

- p) View all the DEM images with DISPLAY Launcher using the default Quantitative palette. When viewing DEM30, click on the Layer Properties button in Composer and change the palette file to be a Qualitative palette. Next, window in on the upper left quadrant. Notice that with a qualitative palette, it is much easier to see the "runways" or "spikes" that travel through the contours. These distortions occur because of the adjacency of contours. When finished viewing the DEMs, view DEM30 as a surface image in ORTHO. Uncheck the Use drape image box and accept all other defaults.

The river valley has a value of 1500 feet. While a 250 foot contour interval is sufficient for interpolating areas of steep slope, a different interval may be necessary to appropriately characterize the valley. Mountains are interspersed with widely connecting valleys which are just as much a part of the mountain system. Failing to represent the more gradual changes in elevation levels accurately is a loss of valuable information for understanding processes occurring in mountains. Understanding the integration of these places in a community's farming system, the nature of travel through mountains, and the relationship of water resources to the physical geography, all require accurate representation of the low sloping valley areas.

- q) Run ORTHO and specify the surface image as DEM90, uncheck the Use drape image box, and click OK. The coarse quality of the DEM at this resolution is apparent. When done, close all the images displayed.

To learn more about the effect of resolution, we will look at the frequency histogram of elevation values in two elevation models. For this, we use HISTO.

- r) Run the module HISTO from the GIS Analysis/Database Query menu and specify DEM90 as the input image. Use the defaults for the remaining options. After the program finishes calculating the histogram, graphical and statistical output of elevation values appear. This histogram illustrates the frequency with which each elevation value occurs. View another histogram for a finer resolution DEM such as DEM60. Then, call up DEM15 in HISTO, again accepting the defaults.

A fine spatial resolution is required to portray the details of a very steep surface. We notice that as the spatial resolution of the DEM images becomes finer, more detailed interpolation occurs between the contours. Notice in the histogram the spikes of regularly-spaced high data columns. These tall spikes represent the cells containing the original contour values. The intermediate bars are the cells between the contours, where elevations have been interpolated. They occur, to different degrees, in *both* images.

The dominance of elevation values of the contours in DEM90 relative to the interpolated values suggests that the resolution is poor, i.e., that the pixel size too greatly generalizes the value of elevation for each pixel space. The spikes in DEM15 are shorter but do not disappear. On the one hand, this is a problem of resolution. If we initialized a finer mesh grid for rasterizing our vector contours, perhaps the dominance of contour values relative to other values would disappear altogether. The elevation surface might become more smooth.

On the other hand, it might not. Using contour intervals as our elevation sample in the first place potentially overrepresents those elevations. An alternative to using contour *lines* is interpolation from sample *points*. Using sample points might reduce the problem of overrepresentation of certain elevations, but such samples are generally not easily available. Interpolation from points also has its own sources of error and distortion and may not necessarily provide a better representation of the true surface.

The problem of interpolating from a sample and the effect of resolution is inherent in any digital terrain model. It is greater in models of very highly contrasted relief, as is clear in our attempts to represent the three-dimensional character of the Himalayas in two-dimensions. One should be conscious of how this may affect the character and quality of data derived from the DEM as demonstrated later in the exercise.

Finally, to alleviate some of the effects of imperfect interpolations, we can run averaging filters across the interpolated image. These serve to smooth the areas of coarseness. However, we must be careful to not eliminate irregular features that are actually part of the rough mountain terrain being represented. As we noted earlier, the choice of interpolating technique has a great influence on the coarseness of the final DTM. Knowing the possible results of a particular algorithm should help determine the value of a smoothing filter. For example, the linear interpolation technique adopted by IDRISI sometimes produces images with sharp angularities for small features. For example, when a mountain peak is represented as a single pixel with few other defining contours, star-like features appear. In cases like this, running the averaging FILTER module on the values of the image one or more times will smooth the DEM (see the IDRISI Help System for a further explanation of the module).

- s) For purposes of illustration, run the module FILTER from the GIS Analysis/Context Operators menu. Retain the default mean filter option. Specify the input image as DEM30 and call the final output image DEMF30. If necessary, change the palette option so both images are displayed with the Qualitative palette. Compare the two images by viewing them side by side. If you look closely, the DEMF30 image appears smoother than the DEM30 image. To further examine the differences between the two images, display the graphic histogram of the values in DEM30 and DEMF30 each with HISTO. Look at the histograms and compare the resulting graphs and statistics.

3. *Explain what happened when you applied the mean filter to the image.*

Choosing Pixel Resolution: An Objective Method ⁴

Contours are a vector representation of height. As demonstrated in the previous example, any application of contours in a raster system artificially gives the contours width. The right choice of spatial resolution for transforming a contour map from a vector to a raster format is therefore crucial in any application involving further uses of the image. A wrong choice of spatial resolution for a contour image can magnify distortions in the interpolation process as well as further distort any additional data derived from the DEM. A coarse grid may not portray the surface in detail, leading to oversimplification. A resolution that is finer than the underlying topographical information may also produce errors in interpolation.

From the outset of any GIS project, it is necessary to determine the resolution that is appropriate for the purpose, and then determine what the available data can provide. One of the most difficult tasks of using GIS in mountain environments is deciding upon the appropriate resolution. Resolution should not be forced upon a project by available data.

The following quantitative method for determining an appropriate cell size for a DTM is provided for consideration. It follows the common-sense notion that a higher resolution grid provides more detail, but that there is a point after which *increasing* resolution (meaning *decreasing* cell size and *increasing* the number of cells used for the represented terrain) provides no real advantage or accuracy. This is rather like the idea behind significant digits in scientific applications. Here, a formula has been provided in an attempt to rationalize and optimize the resolution selection. However, the use of the formula itself is not obligatory in GIS applications. The underlying principle is important however, and reflects the logic necessary in understanding the non-linear relationship between detail and accuracy in topographic analysis and representation.

The formula is based on a theory of information. The theory explains the relationship of the cell size in a DTM to the quantity of information. It suggests that the quantity of information contained in a DTM increases as the cell size of the original contour decreases (see Figure 1). This will continue until a certain pixel size is reached after which no significant increase in information will occur by increasing the resolution. At that critical point, any further reduction in pixel size introduces redundant information. Shannon's Diversity Index (H_T) is derived from a formula using this theory of information (del Barrio et. al. 1992; Turner, 1989). This index represents the information H of a series of N objects which occur in S classes where P_i is the proportion of cells in class i :

$$H_T = - \sum (P_i * \ln P_i)$$

4. This is a direct application of the method developed by del Barrio et al., 1992.

\ln = the natural logarithm

P_i = the proportion of the landscape in each class i (i.e., n_i / N , the frequency of class S_i)

H_T = the total information (in this case, total diversity of the DTM)

The equation can also be solved for a smaller kernel or portion of the image to derive H_I (the local diversity index). In this case, we applied the formula to the data from the contour images created earlier to derive the total diversity and therefore the maximum amount of information available in the image as a whole. Results are presented in Table 1.

Table 1

Summary of Image Characteristics

	Pixel Size	1/ SQRT of Area	H_T (maximum)
DEM90	90	0.0111	2.2330
DEM75	75	0.0133	2.3197
DEM60	60	0.0167	2.3800
DEM45	45	0.0222	2.4307
DEM30	30	0.0333	2.4503
DEM15	15	0.0666	2.4579
DEM7	7	0.1429	2.4519

From Figure 3, we notice that H_T (information) increases as the cell size progressively decreases (also see Table 1). However, when the cells become too small beyond a threshold, the graph begins to level off, indicating the optimum point after which no new information is added as the cell size further decreases. If the series becomes redundant, successive reductions in cell size do not provide anything new, and the increase in total information tends to slow down.

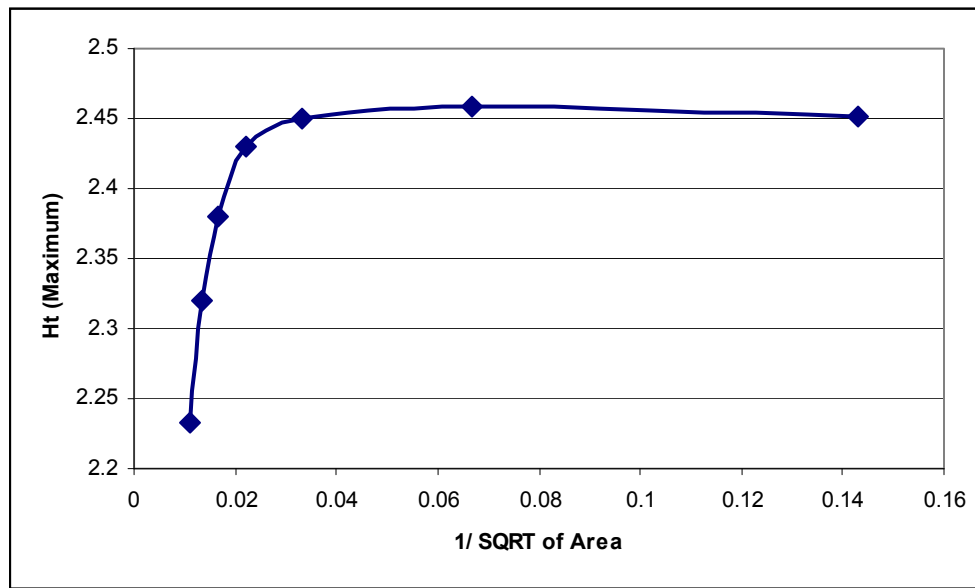


Figure 3: Optimum Information Level

We note from the graph of the global diversity indices that a spatial resolution of 15 meters best represents the optimal pixel size for this contour map. The graph also reveals that the change in the indices between 0.0333 (corresponding with a pixel size of 30 meters) and 0.1492 (representing a spatial resolution of 7 meters) is very small.

Significantly, the information theory from which the formula is derived works from an underlying assumption that the number of grid interpolations should be roughly equivalent to the number of data points. This has an important general implication for the processing of such information as contours. Raster systems such as IDRISI will exhibit a geometric increase in computing time as the pixel size reduces from something such as 60 meters to 30 meters. Increasing resolution not only has a ceiling for the contribution of new information but also pays a price in speed and convenience.

Considering the fact that computation time will increase significantly with any increase in resolution, it will in this case be convenient to choose a spatial resolution lower than 30 meters but higher than 15. This conclusion mirrors the decision when made from empirically examining images with these resolutions earlier in the exercise.

This result applies to the provided data set only and is not a universal prescription. Every data set will have its own resolution threshold. The formula provides a template into which any data set can be inserted. It does point to a general principle concerning the relationship of increasing detail to overall accuracy.

Deriving Data from the DEM

Spatial resolution controls not only the level of detail, representativeness, and processing speed of the DEM but also the accuracy of applications involving the use of the image. In the process of developing a potential hazards map, a health services accessibility map, or any other information to describe or predict processes occurring in mountains, the underlying inaccuracies of the base elevation model affect the accuracy of data derived from the model. This lowers the confidence one can have in using these maps. The question of which spatial resolution to use in working with contours and other relief data is therefore critical to relief surface modeling.

Example: Slope Stability

The DEM can be a source of several thematic maps related to the terrain surface. One such map is slope. Like the DEM, slope and all other thematic maps derived from it are affected by the choice of pixel size. In implementing a GIS, slope stability programs must search for data sets which allow them to address their particular research questions but which they can reasonably afford. A program may use a relatively coarse DEM in order to sort out potential areas of concern for slope stability within a region. But when a project narrows its focus, for example, to the length of a specific ridge, the scale of the required information changes, as does the resolution that best represents the higher levels of detail required.

In this part of the exercise, we will look at the DEM's impact on the calculation of slopes. In IDRISI, slope is directly calculated from elevation using the module SURFACE. This module may also be used in the calculation of aspect or the creation of analytical hillshading models.

- t) Run the module SURFACE from the GIS Analysis/Context Operators menu and choose to calculate slopes (the default option) from the input elevation model DEM60. Call the output slope image SLOPE60 and specify degrees as calculation units (this is the default option). Click OK. A conversion factor needs to be specified because the elevation units are measured in feet and the reference system is in meters. In the Conversion from unspecified to m input box, enter a value of 0.304 (0.304m = 1 foot). Click OK. Repeat the procedure to determine slopes from DEM30 and DEM15 and call the resulting images SLOPE30 and SLOPE15 respectively.
- u) Use HISTO to view the distribution of the slope values of each image. In each case, specify the name of the image for which the histogram should be calculated and click OK to accept the defaults for the remaining choices. If an IDRISI warning message appears indicating that the last class is too small and that the maximum will be reset, click OK. Notice the differences in the maximum value of the images.

Suppose a slope stabilization program needs to isolate areas of slope greater than 45 degrees. How might the different resolutions affect the total area calculated for these slopes? To answer this question, we will reclassify the slope data values into Boolean categories where pixels of slope greater than 45 degrees are given a value of 1 and slopes less than 45 are given a value of zero. We may then calculate the total area of the images that fall in this category, and compare the results. For the reclassification, we will use the RECLASS module.

- v) Run the module RECLASS from the GIS Analysis/Database Query menu. We will be reclassifying an image with the user-defined classification (this is the standard default option for most cases of reclassification). Specify the input file as SLOPE15 and call the output file SLREC15. Assign a new value of 1 for all values ranging from 45 to those just less than 100. Assign a new value of 0 for all values ranging from 0 to those just less than 45. Double-check the values you entered and click OK. Now do the same reclassification to SLOPE30 and SLOPE60. You can look at the results in DISPLAY Launcher with the Qualitative palette to verify that all old values are now in one of the two categories.
- w) Next, for each reclassified image, run the module AREA from the GIS Analysis/Database Query menu and choose the tabular output of the result in hectares. Then copy down the value for category 1 which represents values of a slope greater than 45 degrees, and enter that value in the table below. Do the same for SLREC30 and SLREC60. We calculated the area values of the other images for you. The results are shown below.

slope > 45°	Area
SLREC90	58.24
SLREC60	

SLREC30	
SLREC15	
SLREC7	371.32

4. *How do you account for the differences in the area figures for slopes greater than 45 degrees?*

The remarkable variation poses serious problems for a program's efforts. This demonstrates the high degree of potential variation. Working from inappropriate base resolution, erroneous data may be derived and a program might spend money in the wrong areas without mitigating any of the severe problems posed by unstable slopes. *[Note: Before starting the next exercise, be sure to close all images. In future exercises, users will not be reminded.]*

Conclusion

This exercise briefly presented an overview of some of the issues raised by modeling a three-dimensional surface in two dimensions. We began by examining the contour in its raw form. Next we considered the vector TIN solution of creating a continuous elevation surface for analysis. We then went through the steps of raster DEM construction, showing the places in which resolution and interpolation methods may change the character of the elevation model. Finally, we considered the implications of varied resolution in further processing of DEM data.

Sampling, data structure, interpolation procedure, and resolution all affect the quality, accuracy, and detail of the final DTM. Ultimately, the purpose of the DTM, data availability, accessibility of interpolation procedures, and the analysis of costs and benefits for the production of DTMs, all affect the final decisions made about their use. Finally, models are not truths but simply representations designed to assist the process of deriving new information, decision making, and planning. An awareness of their limitations is necessary. DTMs and the data derived from them will affect decisions in planning and should be created and used with care⁵.

References

- del Barrio, Gabriel, Bernardo, A., and Diez C. J. (1992) "The Choice of Cell Size in Digital Terrain Models: An Objective Method." Paper presented at the Conference on Methods of Hydrologic Comparison. Oxford, U.K. September 29 - October 20.
- Burrough, P.A. (1986) *Principles of Geographic Information Systems for Land Resources Assessment*, Clarendon, Oxford.
- Chen, Z, and J.A. Guevara (1987) "Systematic Selection of Very Important Points (VIP) from Digital Terrain Models for Construction of Digital Triangular Irregular Networks." *Proceedings, AutoCarto 8*, ASPRS/ACSM, Falls Church, VA, 50-56.
- Eastman, J.R., P.A. Kyem, J. Toledano, and W. Jin (1993) *GIS and Decision Making*. United Nations Institute for Training and Research, Explorations in GIS Technology. Volume IV. Geneva: UNITAR.
- Ebner, H. (1987) "Digital Terrain Models for High Mountains." *Mountain Research and Development*, 7 (4): 353-56.
- Lam, N.S. (1983) "Spatial Interpolation Methods: A Review." *The American Cartographer*, 10 (2): 129-49.
- Mark, D.M. (1975) "Computer Analysis of Topography: A Comparison of Terrain Storage Methods." *Geographischer Annaler* 57A: 179-188.

5. Where error in a DTM is known or can be estimated, the uncertainty it introduces may be incorporated into the decision making process. See volume 4 in the UNITAR workbook series: *GIS and Decision Making* for a further exploration of these issues (Eastman, et al., 1993).

- McCullagh, M.J. (1988) "Terrain and Surface Modeling Systems: Theory and Practice." *Photogrammetric Record* 12 (72): 747-79.
- Turner, M.G. (1989) "Landscape Ecology: The Affect of Pattern on Process." *Annual Review Ecological Systems*, 20, 171-97.

Exercise 2: Associations Between Land Use Change and Topography



Introduction

Geographic information systems allow us to document and analyze temporal and spatial patterns of change and assess their potential impact. Of particular concern is the evaluation of the implications of change as it occurs in association with topographic variables, particularly elevation, slope, and aspect. This exercise illustrates GIS techniques that estimate the extent of land use change over time and evaluate the kinds of change occurring for land use policy decision making.

Here, we examine changes in a sub-watershed of the densely populated Jhikhu Khola watershed located 40 kilometers east of Kathmandu, Nepal. We concentrate on the upper part of the watershed called Dhulikhel, which has a high proportion of steep lands. Public perception of large scale and increasing deforestation due to rapid population growth in the middle mountains was very common during the 1970s and early 1980s. Claims went as far as to say that by 1993, there would be few trees left in the hills of Nepal (World Bank, 1978), thus prompting national reforestation programs. Recent studies have since enriched our understanding of the national status of the forest in Nepal. This exercise uses data from one of these studies of this area.¹ For a more detailed analysis of forest changes in the Jhikhu Khola, see Shreier et al. (1994).

We explore land use data for 1972 and 1989 to evaluate the degree to which certain land use practices have expanded while others have declined. We also examine where these major changes have occurred and in what ranges of elevation and slope these changes have been most apparent.

This exercise consists of three sections. In the first part, we analyze dynamics between different land use categories. Using images of the study area from 1972 and 1989, we use spatial crosstabulation to map areas where some land uses have given way to others over these seventeen years. We further measure the total amount of area in which these changes are occurring in order to quantify trends in land use change.

The second section concerns land use dynamics in relation to elevation and slope. We create images showing areas of significant land use change and by incorporating elevation and slope data, determine which land use changes are occurring in which elevation and slope regimes. Results of this analysis allow us to draw some conclusions regarding sustainability of current land use trends.

In the final section, a possible methodology is demonstrated for the use of GIS to generate alternatives to improve resource status in the area. New criteria for the implementation of forest plantations are considered and maps are created to target possible future resource development.

Data layers used in the exercise include:

USE72	land use map for 1972
USE89	land use map for 1989
DH_ELEV	digital elevation model (DEM)
MASK	a Boolean image showing the sub-watershed and the background

Land use maps for 1972 and 1989 for the Dhulikhel sub-watershed were created through photo interpretation of aerial photographs (1:20,000). In addition, the 1989 map was verified through field testing. A DEM was produced using IDRISI's global INTERCON interpolation on contours digitized from a 1:20,000 topographic map for the area.

1. Data for this exercise were made available by Dr. Hans Shreier, Resource Management Sciences, University of British Columbia, Vancouver, British Columbia, Canada and Dr. Margaret Schmidt, Department of Geography, Simon Fraser University, Burnaby, British Columbia. All data we imported from Terrasoft. Contours were re-interpolated in IDRISI.

Land Use Dynamics between 1972 and 1989

- a) Use DISPLAY Launcher from the Display menu to display USE72 with the Quantitative palette and a legend. Display USE72 again with DISPLAY Launcher, but this time with the Qualitative palette and a legend. Compare the palettes of both images by placing them side-by-side. The Qualitative palette differs from the IDRISI default Quantitative palette as it uses color combinations that provide high contrast between the categories. These are designed especially for the visualization of qualitative data categories like landuse, but are of little value when examining quantitative and continuous data sets like elevation or slope. These palettes can also be selected by using the user-defined option before bringing up the display. Notice how many landuse categories there are in the image.

Khet and Bari are two types of agriculture practiced in the sub-watershed. Khet is an intensive irrigated agriculture dominated by rice and is limited to valley floors and level terraces. The irrigation water usually comes from a local creek. Bari is rain fed agriculture dominated by maize. Winter wheat and mustard are also cultivated in both systems. Grasslands are areas which have low (less than ten percent) tree cover. Grazing land often changes into shrubland through natural succession. This sometimes makes distinguishing between them difficult. Both grazing land and shrubland are used as a source of animal feed. The other land categories include urbanized land and transportation. Forests below 1200 m are dominated by tropical and subtropical mixed species, and above 1200 m, by mixed broad-leaved species (Tamrakar et al. 1991). Local farmers traditionally use the forests for diverse purposes: leaves for animal fodder, seeds for medicine and oils, forest litter for animal bedding and as fertilizer in agriculture, and branches and stems as fuelwood (Schmidt, 1991).

Use DISPLAY Launcher again to display USE89 with the Qualitative palette and a legend.

1. *What landuse category exists in the second image but not in the first?*

Afforestation was initiated in Nepal at the beginning of the 1980s in order to regain the forest lost during the late 1950s and 1960s -- a time when government nationalized forests and local groups and communities lost legal control over the use of this resource. The intent of the program was to increase forest cover, produce a source of wood products and stabilize steep slopes. Most of the forest plantations were dominated by chir pine (*Pinus roxburghii*). Pine trees were chosen because of their ease in propagation and management in nurseries and their ability to survive on very degraded soils. It was believed that a secondary forest of broad leaf species would eventually take over the pine forest (Schreier and Brown, 1992).

The objective of this section is to determine the spatial extent of each land use and evaluate the changes in those uses from 1972 to 1989. The first step is to calculate the spatial extent of each land use type in 1972 and 1989. In raster GIS, this is accomplished by having the computer simply calculate the number of pixels in a class and then multiplying by the spatial resolution of the pixels. In this image, the size of each pixel is approximately 100 square meters (an exact figure is calculated in the documentation file which can be accessed by clicking Layer Properties and then View Metadata in the Composer box).

- b) Run the module AREA from the GIS Analysis/Database Query menu on each landuse image and specify that you want tabular output in hectares. Fill out Table 1 with area figures to one decimal place. Ignore category 0 which represents background. Notice that the category values correspond with the legend categories (e.g., the category value "1" corresponds with the legend category "Bari," "2" corresponds with "Khet," and "7" corresponds with "Other"). Percentage figures for 1972 and 1989 already have been calculated for you and are provided for convenience; they will be used in the analysis.

Table 1
Landuse : 1972 and 1989

Landuse type	Area, 1972 (ha)	Area, 1972 (%)	Area, 1989 (ha)	Area, 1989 (%)	1972-1989 change (%)
Bari (terraces)		30.7		39.9	9.2
Khet		6.5		8.4	1.9
Forest		19.6		20.4	0.8
Plantation		-		9.5	9.5
Grassland		20.1		10.1	-10.0
Shrubland		21.9		9.5	-12.4
Others		1.2		2.2	1.0
Total		100.0		100.0	

2. *Which categories have undergone major changes between 1972 and 1989? Has there been any deforestation? Why do you think there was no significant expansion of intensive irrigated agriculture (khet) over these years?*

To understand the character of Bari agriculture and forest cover (in the form of plantations) expansion, we need information on the spatial extent of these changes. Crosstabulation of images is a method that allows the examination of spatial relationships between features in two images. In IDRISI, this operation is called CROSSTAB. This module creates an image or table showing all possible combinations of the categories in the two original images. From this image, it is possible to determine the locations and types of change that have occurred.

- c) Run the module CROSSTAB from the GIS Analysis/Database Query menu. Specify USE72 as the first image in the tabulation, and USE89 as the second. Choose to produce a cross-classification image. Call it USE72_89. Use DISPLAY Launcher to view USE72_89 with the Qualitative palette and a legend. Use the scroll key to view the higher categories.

The column of numbers to the right of the colored boxes shows the category number from the 1972 image. The next column of numbers to the right shows the category from the 1989 image. Each legend category thus shows a combination of landuse types from legends for the 1972 and 1989 images. For example, category 5 reads "5 | 1." This means that all the cells in the crosstabulated image with the value of 5 are areas which were category 5 (grassland) in 1972 but have become category 1 (Bari agriculture) in 1989.

3. *What does category 1 represent in this image?*

The following table is based on the USE72_89 legend. It will facilitate the analysis of land use dynamics. We want to estimate the spatial extent of particular land use changes by calculating their areas.

- d) Run AREA on USE72_89 and ask for tabular output in hectares. (Information can also be output to an ASCII values file and then imported into a spreadsheet program.) Record the area figures from the screen in the following table. The first three columns correspond to those of the legend in the USE72_89 image. For ease of reference, we have replaced the numeric values of categories with the category name they represent. Notice that categories like number 2 (Bari in 1972 and 1989) reflect areas of no change. Judging from both this table and from Table 1, it is apparent that the areas of significant change are those where Bari agriculture and plantations have expanded and where grass and shrubland have declined.

Table 2
Areas of Landuse Change

Legend Number	Landuse, 1972	Landuse, 1989	Area, ha	% change, 1972-89
2	Bari	Bari		
3	Khet	Bari		
4	Forest	Bari		
5	Grassland	Bari		
6	Shrubland	Bari		
7	Others	Bari		
8	Bari	Khet		
9	Khet	Khet		
10	Forest	Khet		
11	Grassland	Khet		
12	Shrubland	Khet		
13	Others	Khet		
14	Bari	Forest		
15	Khet	Forest		
16	Forest	Forest		
17	Grassland	Forest		
18	Shrubland	Forest		
19	Others	Forest		
20	Bari	Plantation		
21	Khet	Plantation		
22	Forest	Plantation		
23	Grassland	Plantation		
24	Shrubland	Plantation		
25	Others	Plantation		

26	Bari	Grassland		
27	Khet	Grassland		
28	Forest	Grassland		
29	Grassland	Grassland		
30	Shrubland	Grassland		
31	Others	Grassland		
32	Bari	Shrubland		
33	Khet	Shrubland		
34	Forest	Shrubland		
35	Grassland	Shrubland		
36	Shrubland	Shrubland		
37	Others	Shrubland		

To calculate percent of land use type change, divide the area figures of Table 2 by the 1972 area figures for the original land use category in Table 1. Multiply the result by 100 to obtain the percentage. This represents the percentage of the 1972 land use category which has changed into the new category. For example, category 5 shows all the areas which were grassland in 1972 and have become areas of Bari agriculture in 1989. This category covers 69.2 hectares. By dividing by the original area of grassland coverage, 255.3, we see that 27.1% of the original grassland has become Bari agriculture.

4. *Bari agriculture expanded at the expense of which categories? And forest cover (in the form of plantations)? How can you explain that almost 22% of 1972 forest cover was categorized as forest plantations in 1989?*

There are a variety of other methods which could be used to explore trends in the data. For example, CROSSTAB can also provide summary statistics showing the correlation between the old and new images. A Kappa index of agreement can be derived to quantify the extent to which the two images are different and thus, the extent to which change has occurred throughout the region. For more on the use of this kind of analysis, see Eastman, McKendry, and Fulk (1994). For now, we move on to the next set of questions of concern to this exercise.

So far, we have examined and compared two land use images of the region and identified land uses which have significantly expanded or declined between 1972 and 1989. By crosstabulating the images, we have further determined the degree to which particular land uses have given ways to others. Land use changes in mountain ecosystems often have a vertical as well as horizontal component; some land use changes are often concentrated in areas of particular slope and elevation. It is necessary then, to introduce topographic data into our analysis to fully understand the character of change in this region.

Land Use Change in Relation to Elevation and Slope

In order to gain a better understanding of the implications of forest and agricultural expansion and grazing and shrubland decline, it is necessary to explore these gains and losses in relation to site topography. In particular, we want to identify such changes in areas of critical high slope and elevation.

To proceed with this analysis, we will explore the elevation and slope characteristics of the region and determine where

areas of high slope and elevation occur. We begin by redefining the continuous surfaces of elevation and slope values into discrete classes for ease of analysis. We then compare our land use images relative to these classes, determining which ranges of elevation and slope are the sight of particular land use trends.

Elevation

- e) Use DISPLAY Launcher to view DH_ELEV, the digital elevation model for the sub-watershed, with the default Quantitative palette. This model was produced from a digitized contour map by rasterizing the vector lines and using INTERCON as the interpolation procedure. The resulting raw DEM was smoothed using the FILTER module. Use Cursor Inquiry Mode to explore the range of elevations (the "z" value) in the model. [Note that the elevation units are in meters.]

5. *What is the elevation range in the sub-watershed? What is another way to determine this information?*

- f) Run HISTO to look at the frequency distribution of the elevations in the DH_ELEV image. Be sure to select a graphic output and use all other defaults.

Notice that there are regularly repeated peaks in the histogram. These peaks represent the points corresponding to the contour lines of the digitized map. Distances between them in the histogram show contour intervals of the original map.

6. *What is the approximate minimum elevation in the model? What is the contour interval of the map that was digitized?*

To use elevation data in the analysis, it is necessary to reclassify the continuous elevation surface into discrete categories, each representing a range of elevations. Therefore, the dominance of the contour interval in the data set may not be a significant issue.

- g) To create these categories, run RECLASS and specify the input file as DH_ELEV and call the output file ELCLASS. Assign a new value to the values in the following ranges:

	lower	upper
new value	old value	old value
0	0	800
1	800	1000
2	1000	1200
3	1200	1400
4	1400	1600
5	1600	1800

When you have finished assigning values, check that you have input the correct values and then click on OK.

7. *Why did we choose 800 m as the upper limit of the first class (new value of 0)? Could we have chosen 1000 m as the upper limit to this class? What would have happened in this case?*

Open Metadata in IDRISI Explorer, select the ELCLASS file and double-click on the Categories option to add legend categories. Add the first caption as "800 - 1000m." Do the same for legend categories 2 through 5, updating the categories with the ranges assigned from the table above. When done, save the changes. Use DISPLAY Launcher to view ELCLASS with the Qualitative palette and a legend.

For the later analysis of land use dynamics, specifically the amount of change occurring within each elevation category by land use type, we will need to know the area of the sub-watershed that falls within each of these elevation classes.

- h) Run AREA on ELCLASS, specifying tabular output in hectares. Record the area of each elevation regime. Remember that category 0 is the background of the image and does not represent an elevation category.

8. *What elevation range is predominant in the sub-watershed ?*

Slope

We have now created elevation classes from our DEM and have derived area measurements of the coverage of each range of elevation. We next derive a slope surface from the DEM and similarly divide it into range classes. We then compare the two images, determining the relationship between elevation and slope in the region by finding the elevation ranges within which high slopes occur.

- i) Before calculating slopes from DH_ELEV, open Metadata to examine the file contents. In particular, look at the flag value and flag definition fields.

In this case, 0 is flagged as a special value. Since it is designated as "background," it represents any pixel beyond the extent of the study area. The SURFACE module in IDRISI is designed so that any pixel with a value of 0 will be left out of slope and aspect calculations and so prevents interpolation beyond the study area. This avoids the creation of extreme or irregular values at the study area's edge.

- j) Run the module SURFACE and specify the input elevation model as DH_ELEV and call the output image SLOPERAW. Specify percent as the slope measurement units and input a conversion factor of 1. If prompted that data flagged as missing or background values will be given a slope of 0, click OK. Examine the output in DISPLAY Launcher using the default palette.

Slope gradients in percent express the relative change in height for any distance traveled in a horizontal direction. Due to the algorithm structure of SURFACE, there are some linear strikes near the flat areas in the sub-watershed. Also, the impact of the dominance of the contour interval relative to resolution becomes apparent in the slope image.

- k) To smooth the distribution of slope values, run the module FILTER from the GIS Analysis/Context Operators menu. Retain the default mean filter. Specify the input image as SLOPERAW and call the filtered image SLOPEF. (Note that a comparison of the frequency histograms illustrates the impact of the filter.) Use DISPLAY Launcher to view SLOPEF with the Quantitative palette.

- l) Again, we want to work with categories of slope range rather than the continuous raw slope values. Run RECLASS to reclassify the image SLOPEF into a new image called SLOPEREC. Assign the new values as follows:

class 1 : 0 - 5 %

class 2 : 5 - 20 %

class 3 : 20 - 35 %

class 4 : 35 - 50 %

class 5 : > 50% (use an arbitrarily high value, for example 999, to include all other values)

Look at SLOPEREC with DISPLAY Launcher using the Qualitative palette and a legend.

As you can see, background cells are classified as class 1. This is because both flat (no slope) areas in the study region and

background areas falling outside the study region were given values of 0 in the original slope image. The computer had no way of distinguishing these distinct areas during reclassification. To exclude the background areas from class 1, a mask image of the study area is needed. We use this mask to change background values in SLOPEREC from 1 to 0, without altering cells belonging to category 1 that fall within the study region.

- m) Run OVERLAY from the GIS Analysis/Database Query menu. Specify SLOPEREC as the first image, MASK as the second, and call the output image SLOPES. Select the multiply overlay option (First * Second). MASK is an image showing the watershed as a value of 1 and the background as 0. This will simply multiply all cells within the study area by 1, leaving them unchanged, while multiplying the cells in the background by 0, removing them from category 1. You may wish to redisplay this image with a Qualitative palette. Next, go to Metadata and select the SLOPES file, then double-click on the Categories option to add legend categories. Enter slope categories as done for the elevation legend previously where category 1 will read "0 - 5%" and so on. Use DISPLAY Launcher to view SLOPES with the Qualitative palette and a legend.

Change in Relation to Elevation and Slopes

Now that we have maps depicting classes of slope and elevation, we can analyze land use change relative to these topographic conditions. To make this analysis easier, we first create four Boolean images each showing separately the areas of forest cover expansion, agricultural expansion, grassland loss, and shrubland loss. Since the plantations make up the bulk of the expansion of forest cover, we will limit our analysis to plantations within the forest cover image.

To create an image showing Bari agriculture expansion, we need to create a Boolean image showing only these areas. To do this, we simply reclassify the categories representing this expansion in the image USE72_89 to have a value of 1 and change all other categories to 0. For this purpose, we will use a combination of IDRISI modules to perform reclassification called Edit/ASSIGN. ASSIGN gives new values to a set of integer data values based on the information contained in a values file.

- n) First, we need to create an attribute values file with new assignments for the data values. Run Edit from the Data Entry menu. Type in two columns of numbers as follows, making sure to leave a space between the two values:

2 1

3 1

4 1

5 1

6 1

7 1

This means that the old categories 2 through 7 from USE72_89 will be reassigned to a new category 1. These are all the categories showing areas into which Bari agriculture has expanded. You can confirm the meaning of these by looking at Table 2. Data values that are not listed in the file will automatically be assigned a new value of 0.

When finished, click the Save button. Specify BARIEXP for the filename and in the Save as Type box, choose Attribute values file. Click the Save button again. You will be asked whether you want to save the data as real or integer data type. Click OK to accept the default of integer data type and close the Edit dialog box. Now run ASSIGN from the Data Entry menu. Specify USE72_89 as the feature definition file, BARIEXP as the attribute values file, and call the new output image BARIEXP. Since the BARIEXP files have different extensions in their filenames, one will not overwrite the other. Use DISPLAY Launcher with the Qualitative palette to look at

BARIEXP. It should be a Boolean image showing areas of Bari agriculture expansion with a value of 1.

- o) Repeat the same procedure three more times and create maps of plantation expansion (PLANTEXP), grassland loss (GRASLOSS), and shrubland loss (SHRULOSS). Use Edit to create three attribute values files. Again use category numbers from Table 2 in the left hand column with the value 1 in the right hand column. For plantation expansion, use the categories which show plantation areas in 1989 which were formerly other land cover. In the case of grass and shrubland loss, use the categories of USE72_89 which show grass or shrubland coverage in 1972 but a different landuse in 1989. Use ASSIGN to apply these values files to USE72_89.

We now have four Boolean images showing the areas where significant changes have occurred between 1972 and 1989. These can now be used in combination with slope and elevation data. One method for this analysis involves the extraction of mean pixel values in areas of change. In IDRISI, we use the module EXTRACT.

- p) Run the module EXTRACT from the GIS Analysis/Database Query menu. Specify BARIEXP as the feature definition image and DH_ELEV as the image to be processed. Select the summary type to be the average value and the output type to be tabular. The simple table that appears shows the mean elevation values for pixels with the value 0, which did not experience a change into Bari agriculture, and value 1, which did.

This kind of analysis can be run to calculate the mean slope or elevation values for any area of change. Such information may be sufficient for some kinds of analysis and planning. Feel free to experiment with EXTRACT on various combinations of data. More detailed summaries are also possible however, by overlaying input images.

We can also overlay the images in various combinations to map areas of change in different slope and elevation and slope regimes. We begin with elevation classes. To do this, we combine these images through simple map algebra with the classified elevation model to create four new images showing the elevation classes in each of the areas of land use change.

- q) Run OVERLAY and specify BARIEXP as the first image, ELCLASS as the second, and call the resulting image BARELEV. Choose to multiply the first and second images (First * Second). Display BARELEV in DISPLAY Launcher using the Qualitative palette and a legend. What we see here are the possible combinations of the two images. Cells showing a value of 0 are areas which did not experience a change into Bari agriculture. The remaining cells have received their new value based upon their elevation. Numbers in the legend correspond to the elevation classes from the ELCLASS image.

9. *Which elevation class is missing? What does this mean?*

- r) Repeat the same procedure for the other three images, calling the resulting images PLAELEV, GRASELEV, and SHRUELEV respectively. Look at the images and visually estimate in which elevation class the expansion or loss of a landuse type primarily has occurred. To examine the situation in greater detail, run AREA on each of the four images (BARELEV, PLAELEV, GRASELEV, and SHRUELEV), indicating that you want tabular output in hectares. Write down the results in Table 4. The percentage of cells falling into varying elevations can be calculated for each image.

10. *Where did most of agricultural expansion occur -- at lower or higher elevations? At what elevations did most of the grazing land and shrubland loss take place? What is the total area of agricultural expansion versus the total area of forest expansion in the upper elevations (above 1200 m)?*

Table 4

Landuse dynamics in relation to elevation

Elev. class	Bari expansion		Forest expansion		Grassland loss		Shrubland loss	
	ha	%	ha	%	ha	%	ha	%
800- 1000 m								
1000- 1200 m								
1200- 1400 m								
1400- 1600 m								
1600- 1800 m								
Total								

- s) Repeat the procedure for slopes. Run OVERLAY to multiply the images of expansion areas (BARIEXP, PLAN-TEXP, GRASLOSS, and SHRULOSS) with the SLOPES image and call the output images BARSLOPE, PLASLOPE, GRASLOPE, and SHSLOPE respectively. Then run AREA on the resulting images. Fill in Table 5 with the results you obtain from each analysis.

Table 5

Landuse dynamics in relation to slope

Slope Class	Bari expansion		Forest expansion		Grassland loss		Shrubland loss	
	ha	%	ha	%	ha	%	ha	%
0-5% Slope								
5-20% Slope								
20-35% Slope								
35-50% Slope								
> 50% Slope								
Total								

11. *What percentage of agricultural expansion is occurring in slopes higher than 50%? What percentage of forest plantation is occurring in slopes of less than 50%? Grass and shrub coverage is declining in what slope regimes? Based upon the trends shown in Tables 4 and 5, what can be suggested about the sustainability of current land use trends?*

Agriculture is moving upslope to steeper and more marginal areas, removing the soil-preserving grass and shrub covers in its wake. Forest plantations, aimed in part at soil and slope stabilization, are being initiated in lower slope regimes as often as higher ones, where they would likely do more good.

In addition to the marginalization of agriculture and the mistargeting of plantation development, another consequence of land use change over these 17 years is an increasing shortage of animal feed resulting from the loss of graze and browse in shrublands to forest plantations (pine needles are not palatable for animals). Soil fertility is another resource problem aggravated by changes in landuse practices. According to a soil survey (Schreier and Brown, 1992), soil acidity is significantly greater at higher elevations than at lower elevations. Pine forests produce an acidic, low-nutrient litter which is collected during the dry season to use as animal bedding. Later it becomes the major component of the compost-manure input into dryland agriculture, leading to further acidification of already acidic soils of higher elevations. Due to critical feed shortages, all palatable leaf growth in the understory of pine plantations is collected to maintain feed supply, which does not allow broad leaf forest to recover as a secondary forest and eventually replace pine plantations.

Based on these observations, we can conclude that the overall resource condition in the sub-watershed has not necessarily improved with the gain of forest cover. There are suggestions that the number of individual trees on private land has increased and that this might reduce the problem somewhat, but this cannot be examined with GIS due to scale and resolution problems. Differential GPS field surveys might be used to document such effects. Beyond its use for critical analysis of trends in land use change and policy, GIS might in this case be employed to model better programs and target areas for resource development. We conclude this exercise by considering a simple example of this kind of modeling in the case study area.

An Alternative Forest Management Strategy

GIS can play an important role in developing alternative forest management strategies to improve the resource status in the area. We will illustrate how to develop an alternative afforestation plan using GIS. For purposes of illustration, we create a map showing major microclimatic zones in the area based on slope, aspect and elevation conditions only. Each of these zones will represent optimum ecological conditions for different native fodder tree species. These trees can stabilize the most critical slopes, provide animal fodder and fuelwood, and improve soil fertility (Schreier and Brown, 1992; Shreier et al., 1994). We can assign appropriate species to each microclimatic class and create a potential ecological plantation map.

First, we need to divide the sub-watershed into two elevation zones (above and below 1200 m). This elevation corresponds to a threshold in climate and natural vegetation cover.

- t) Run RECLASS on the input file DH_ELEV to create the output file named EL1200. Use the default user-defined classification type. Assign a new value of 1 to all elevation classes from 1 m to those less than 1200 m, and a new value of 2 to classes above 1200 m.

Next, we have to create an image showing dominantly north- and dominantly south-facing slopes. This subdivision is critical because south-facing slopes are significantly drier than north-facing slopes. The direction in which a slope is facing is described as *aspect*. Aspect is calculated as an azimuth; output values range from 0 to 360 degrees (clockwise around the points of the compass). 0 and 360 degrees therefore represent slopes facing due north, and 180 degree values represent due south.

- u) In addition to defining slopes, the SURFACE module can also calculate aspect. Run SURFACE and select to calculate the aspect. Specify the elevation model as DH_ELEV and call the output image ASPECT. Enter a conversion factor of 1 and click on OK. If prompted that data will be flagged, click on OK. Use DISPLAY Launcher with the default Quantitative palette to look at ASPECT. Then view the metadata for ASPECT to find the minimum and maximum data values. A value of -1 represents the aspect of flat areas and background.
- v) Run RECLASS on ASPECT to create a file NSASPECT in order to show dominantly north- and dominantly south-facing slopes as separate categories. Assign new values as follows:

new value	aspect (deg)	orientation
1.0	0 - 45	N - NE
0.0	45 - 135	NE - E - SE
2.0	135 - 225	SE - S - SW
0.0	225 - 315	SW - W - NW
1.0	315 - 360	NW - N

Now display NSASPECT with the Qualitative palette.

12. *Which direction do the majority of slopes face: north or south?*

The next step is to combine both of these images to create a map of microclimatic conditions by elevation and aspect.

- w) Run the module CROSSTAB using the Hard Classification option on EL1200 and NSASPECT. Choose to produce a cross-classification image only and call it CLASSES. Look at the image with DISPLAY Launcher using the Qualitative palette and a legend. Notice that there are more than four classes in the image. Write down the legend numbers for the categories of our interest. Be careful when interpreting the legend. Remember that in

EL1200, 1 represents elevations below 1200 m and 2 represents elevations above 1200 m. In NSASPECT, 1 represents north-facing slopes and 2 south-facing slopes.

- x) Using Edit, create a values file called ZONES that assigns new numbers from 1 to 4 to the four microclimatic zones of interest as follows:

Altitude (Category)	Aspect Category	New Category
< 1200 meters (1)	North-Facing (1)	1
< 1200 meters (1)	South-Facing (2)	2
> 1200 meters (2)	North-Facing (1)	3
> 1200 meters (2)	South-Facing (2)	4

All the other categories (including flat lands) will default to 0. Save the new attribute values file as integer data type. Run ASSIGN on the feature definition image CLASSES, using ZONES as the attribute values file as well as the resulting output file. Use DISPLAY Launcher to view ZONES with a Qualitative palette and a legend.

Another possible topographic condition to account for in creating a new afforestation plan is slope. Due to the importance of agriculture, all land with slopes <35% shall be reserved for agricultural development and excluded from evaluation for forestry.

- y) Use Metadata to find the range of slope values corresponding to slopes > 35% in the image SLOPEF. Run RECLASS on SLOPEF to create a new image SLOPE35. Assign slopes less than 35% a value of 0 and slopes 35% or greater a value of 1.
- z) Run OVERLAY to multiply ZONES and SLOPE35 to produce the final map of microclimatic conditions. Call it FINAL. Display FINAL using the Qualitative palette and a legend. This map shows four microclimatic zones suitable for different tree species considered for afforestation.²

Another component in the development of an alternative afforestation plan is to create a library of native fodder trees where optimum microclimatic and ecological growing conditions are documented for each tree specie. A list of the species for the watershed was developed by the Mountain Resource Management Project (Schreier et al., 1994). Each specie can be assigned to different locations depending on its appropriate microclimatic class thus producing a potential ecological plantation map. Here, we have demonstrated only the application of topographic variables. Certainly the incorporation of additional information on soils, weather, and drainage conditions will refine the process before integrating such results with social and economic criteria for planting in these areas.

In summary, GIS can be used to track the trajectory of change, mapping areas where certain land uses are giving way to others. In mountainous conditions, these trajectories of change can be described and analyzed based on their trends in different elevation and slope regimes. Here, an overall expansion of agriculture and forest areas in Nepal, which might perhaps otherwise be interpreted as a positive trend, can be shown to be potentially unsustainable when considered in light of topographic variables. GIS allows the analysis of different components and dimensions of change in mountain ecosystems. Finally, GIS allows the formation of alternative programs by incorporating topographic variables into suitability analysis and by creating agro-climatic images for use in locally appropriate development planning.

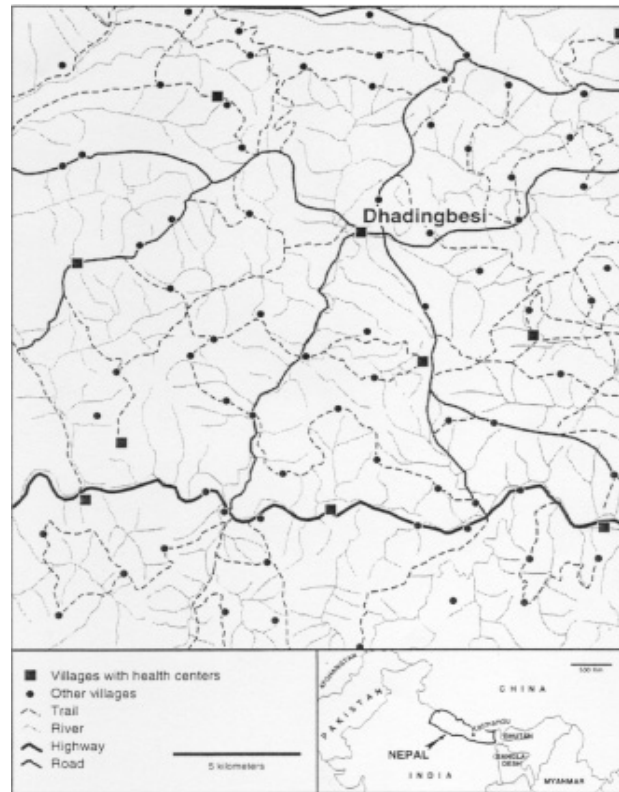
References

Eastman, J., McKendry, J., and Fulk, M. (1994). *Explorations in Geographic information Systems Technology Volume I: Change and*

2. Additional climate data can be used to quantify these microclimatic conditions by employing cooling rates for north and south facing slopes.

- Time Series Analysis*, second edition. Geneva: United Nations Institute for Training and Research.
- Schmidt, M. (1991). "An Evaluation of the Forest Resources and Forest Soil Fertility of a Mountain Watershed in Nepal Using GIS Techniques" *Soil Fertility and Erosion Issues in the Middle Mountains of Nepal*. Workshop Proceedings, April 22-25, 1991; pp. 267-275.
- Schreier, H., Brown, S., Schmidt, M., Shah, P., Shrestha, Nakarmi, G., Subba, K. and Wymann, S. (1994). "Gaining Forest but Losing Ground: A GIS Evaluation in a Himalayan Watershed" *Environmental Management*, 18(1): 139-150.
- Schreier, H. and Brown, S. (1992). "GIS Approaches to Resolve Resource Conflicts in Himalayas" *Geo Info Systems*, October 1992, pp. 52-56.
- Tamrakar, R., Jabegu, K. and Shrestha, B. (1991). "Land Use Changes in the Jhikhu Khola Watershed Area" *Soil Fertility and Erosion Issues in the Middle Mountains of Nepal*. Workshop Proceedings, April 22-25, 1991; pp. 201-207.
- World Bank, (1978). *Nepal Staff Project Report and Appraisal of the Community Forestry Development and Training Project*. Document of the World Bank, Washington, D.C.

Exercise 3: Modeling Access to Health Care: Anisotropic Cost Distance



Introduction

This exercise explores the problems involved in modeling movement in a mountainous environment. In particular, it addresses the complexity of assessing the costs of travel in different directions through an area of high relief. Using data from the Nepalese Himalaya, the exercise seeks to discover the least cost paths between villages and health care centers in the rural mountains.

The exercise is divided into three parts. We begin with a discussion of distance analysis in GIS, introducing the concepts of cost distance and anisotropic cost distance. Next, we introduce the data set for the exercise, a large set of images including factors such as relief, land cover, roads, settlements, and bridges. The variables used are hypothetical, but demonstrate the kinds of variables that might be put to use in the Nepalese context. Although an analysis of this kind has not been done for this region before, it reflects the kinds of questions being pursued by researchers in the region. In the third section, we use the concept of anisotropic cost distance to model inter-village movement in two directions, producing a set of cost distance surfaces and least-cost pathways through the complex terrain.

Distance

Distance analysis is a standard feature in raster GIS. Simply, it involves the calculation of a distance surface from a given location. This source location is generally a feature or point in a raster image (the source image). Distance is calculated by adding the pixel widths (resolution) away from the source. One can imagine walking away from the source pixel with an odometer and steadily adding the accumulating distance. New values are assigned to pixels in the resulting distance surface based upon their distance from the source. Each pixel will now have a value (in the units of the original image resolution) defined by the distance of that pixel away from the source. If there is more than one source, the distance image will give values to pixels reflecting their distance to the nearest source.

This is simple Euclidean distance. Distance is accumulated evenly in any direction and calculated "as the crow flies." The terrain through which the distance is calculated is understood to be homogeneous and unvarying. Euclidean distance surfaces are smooth, with low values around the source, evenly rising to the high values on the margins. This method of distance analysis is effective, simple, and frequently used. It can be used for example, to create buffers and distance data layers for suitability maps. For the calculation of movement through real-world terrain however, simple distance analysis has some limitations.

In mountain environments for example, the problem of slope length can make simple distance analysis problematic. Distance between two points over a sloping surface is longer than over a flat path between them (Figure 1). This problem is a complex geometrical one, which will not be dealt with here.

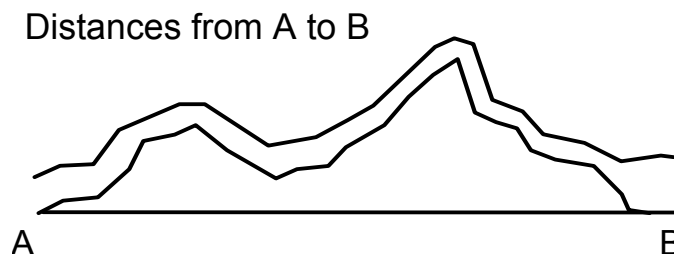


Figure 1: Distance in Mountains

Another, perhaps more critical problem experienced in mountain environments and elsewhere, is that of friction which hampers or impedes movement. Consideration of this problem involves the calculation of *cost distance*.

Cost Distance

Movement across varied landscapes, containing different kinds of land use and land cover, is more complex than movement across a homogeneous plain. Often movement in these kinds of environments involves the expenditure of different amounts of energy to cross different types of land cover. For example, walking across lakes or rivers, may be considerably more difficult or expensive than walking through an open field.

To model this kind of variation in GIS, cost distance analysis is employed. Here, distance is calculated with the incorporation of a *friction surface*. This surface is a raster image in which the pixel values reflect the level of difficulty or expense required to move through them. These surfaces can be created by assigning new friction values to existing land cover images. When computing cost distance, a *distance surface* is again calculated away from a *source*. Here however, distance across any pixel is multiplied by a friction value given in the *friction surface*. The values in the output *cost distance surface* will reflect the distance of each pixel away from the source in units of "cost" (calories, time, dollars, arbitrary relative units, etc.) determined by the values in the input friction surface.

In this case, distance is non-Euclidean, and the resulting surface will not be a smooth one. Distances will rise away from the source, but in an uneven pattern, accumulating costs differently in any direction. This form of analysis is good for modeling movement across real-world landscapes. Using a cost distance surface, a *least cost path* can be calculated to show the most direct route from the source to a given location. This analysis requires an additional step. The path must be calculated using the *cost distance surface* and an image containing a *target* feature. The output image will show the least cost path connecting the two. Problems occur when this form of analysis is used in mountain environments. For movement through mountains, the obvious friction is slope. As slope increases, so does the cost of travel. However, the effect of slope on travel depends upon the direction of that travel. Movement upslope is considerably more difficult than movement downslope. A simple cost distance analysis cannot incorporate the direction of travel (Figure 2).

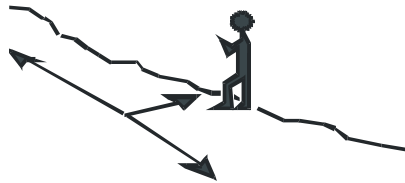


Figure 2: Slope and Direction

Anisotropic Cost Distance

Cost distance frictions like slope, which depend upon the direction of movement, are referred to as *anisotropic*. These kinds of frictions are not limited to mountain environments and include such earth features as ocean currents and winds. Modeling them requires the input of new information into the analysis. Specifically, an anisotropic friction surface will possess *magnitude* and *direction* and must be modeled using a *function* which relates the magnitude to the cost of travel. The magnitude and direction must be incorporated as separate images.

Remaining with the example of slopes, the *magnitude* of slope would be included in a slope image, with degrees of slope being a measure of a slopes friction "intensity"; as slope degrees increase, so does friction. The *direction* of slope would not be included in such an image. A separate image, showing slope direction (aspect) is necessary as well. Finally, a *function* which relates the slope magnitude and direction to cost of travel is required. The function should be designed such that it will transform slope into increasing friction when moving uphill, and a zero friction when moving laterally along a slope. The function could be further designed to create increasing friction for steep downhill movement as well. Magnitude, direction, and their relating function would be used along with a *source* image to create a cost distance surface across a mountainous area.

It is also possible to analyze cost distance using both isotropic and anisotropic costs. Slopes (an anisotropic cost) could be used along with land cover (an isotropic cost) to create a complex surface. In a final step, a least cost pathway could be calculated to any *target* feature. It is this form of modeling that we conduct in this exercise.

Anisotropic Cost Distance in the Mountains of Nepal

This exercise addresses some problems of applying GIS in environmental and health management issues in the Middle Mountain area of Nepal. Broadly, the purpose is to demonstrate the potential of a raster-based GIS to bring together environmental and socioeconomic data in a geographic analysis of an area under question. The specific task is to model the costs of traveling through a mountain environment. A second task is to demonstrate the application of a raster GIS for assessment of distribution and the consequent reallocation of social services in a mountain area based on this cost-distance analysis. Appendix A pursues the latter of the two tasks in detail.

In the exercise, we assess the distribution of health centers based on accessibility of these centers from villages where health services are absent. We assume that people travel by foot (to account for maximum possible costs) and estimate the existing distribution by finding least cost routes from each village to a closest health center and back. We model the difficulty of travel based upon isotropic and anisotropic costs. Isotropic costs include those incurred by land cover and roads which are experienced similarly no matter what the direction of travel might be. Anisotropic costs, in this case slope, are magnified or reduced depending upon the direction of movement. By combining these features in our calculation, we may model the real distances and mark the likely paths tread by villagers in the region.

We will create the cost distance surface for the area using the IDRISI module VARCOST. To conduct the analysis, we require the following inputs:

- 1) A Source Image (or set of images). An image showing villages in the area.
- 2) An Isotropic Friction Surface. An image combining information about the location of land covers, bridges, and roads with assigned friction values for movement through each.
- 3) An Anisotropic Magnitude Image. A slope image derived from a DEM, rescaled to reflect friction values.
- 4) An Anisotropic Direction Image. A variant of an aspect image derived from a DEM.
- 5) A Target Image (or set of images). An image showing the location of health care centers or other villages.

Database

The study area is the central portion of Dhading District, located in the Middle Mountain region (Bagmati Zone) of Nepal.¹ The following data are used:

DHADDEM	DEM interpolated from contours digitized at 1:50,000
ROADS	roads and trails digitized at 1:125,000
BRIDGES	suspension bridges digitized at 1:125,000
RIVERS	rivers digitized at 1:50,000
STTLMNTS	villages and health centers digitized at 1:125,000

1. The MENRIS/ICIMOD staff has developed a vector-based case study (using Euclidean distance) for the Lalitpur District, Nepal. They have contributed the data for developing this raster-based exercise. The original data set was digitized by the MENRIS staff of ICIMOD, Nepal.

LANDUSE 1989 landuse rasterized from polygons digitized at 1:50,000

For the purpose of the exercise, we edited most files in the lab and added features using different sources and techniques. For this reason, the location of villages, rivers, and roads in many cases only approximates the location in the real world. These data should therefore be used for educational purposes only. A survey of the data follows.

Digital Elevation Model (DEM)

The elevation data for the study area are stored in the raster file DHADDEM. A DEM with a cell resolution of 10 meters was interpolated in IDRISI from digitized contours. The result was generalized to a larger cell resolution of 60 meters. The coordinate system uses meters and the elevation values, originally in feet, were converted to meters.

a) Use Metadata to view the file DHADDEM.

1. *What is the range of elevation values in DHADDEM?*

2. *What is the size of the study area in kilometers?*

Display DHADDEM with the default palette. The surface relief changes rapidly within this area and in order to travel through it in any direction, one would have to cross fairly high mountains, large river valleys, hills, and numerous small rivers scattered throughout.

Rivers

b) Now display the vector file RIVERS using the Qualitative symbol file.

As is visible, the Dhading area is very rich in large and small rivers. The rivers are divided into three classes depending upon their size. The largest river, Trisuli, crosses the area in a East-West direction. Some of its big tributaries flow from North to South. There are many small mountain rivers running in different directions. Though small, they become dangerous torrents during the monsoon season.

Land Cover

The image file LANDUSE contains information about types of land cover encountered in the area.²

c) Display LANDUSE with the Qualitative palette and a legend. Five types of landuse prevail in the region: valley agriculture, hillslope agriculture, nonagricultural land (including river banks, urban territories, etc.), grazing land, and forest and shrub.

In the previous exercise analyzing land use change in nearby Dhulikhel, we determined at which elevations these landuse types tend to occur. For this study area, urban land uses, large riverbanks, and valley agriculture usually are found in lower parts of the area, while grazing land and hillslope agriculture exist in higher elevations. Forested areas are spread throughout a wide elevation range.

2. Rasterized from files digitized by the MENRIS staff of ICIMOD from the Land Resource Mapping Project (LRMP) Map Series, 1986.

Villages

The location of villages in the Dhading District is recorded in the point file STTLMNTS. Every village has its own unique ID number ranging from 1 to 81.³

- d) Use DISPLAY Launcher to display the vector file STTLMNTS with the standard default symbol file. This image shows the location of villages in the region.⁴

Next, run Edit and choose Open from its File menu and in the Files of Type box, select Attribute Values File. Open the file named HEALTH.AVL.

The file HEALTH contains information on the availability of health services in these villages. Four types of health facilities are available in this area coded accordingly:

- 1) villages without any health service,
- 2) ayurvedic clinics that provide consulting based on indigenous medicine,
- 3) health posts with temporary medical personnel with some very basic health services available,
- 4) one health clinic in the largest town, Dhadingbesi, with permanent paramedic personnel and temporary presence of a doctor.

Roads, Trails, and Bridges

The location of roads and trails is recorded in the vector file ROADS.

- e) Use DISPLAY Launcher to view the vector file ROADS with the Qualitative symbol file.

The district does not have many good roads. The only highway (paved road, light green) crosses it in an East-West direction along the major river. This road is functional all year and is suitable for any kind of transportation. The next important road (dirt road, light blue) intersects the highway from the North. It is accessible for vehicles and usable for walking trips from October to April. Every year monsoon winds and rain disrupt communication and transportation for a period of five months from May to September.

All other features displayed are trails of varying size suitable only for walking. The only major trail (light brown) that leads to the East-West is located in the Northeast part of the region. Some of the remaining trails are a little larger and of greater importance (red) than others (purple) for communication between villages. The climatic conditions, however, are such that none of the trails in the northern part of the area are functional during the monsoon season (May-September) and winter (December-January).

- f) A vector file BRIDGES contains 8 bridges in the area. Display it with the default symbol file. You may also want to use Add Layer in the Composer menu to overlay the vector file RIVERS with the Qualitative symbol file.

We assume that besides these digitized bridges, there exist other ways (bridges, ferries, etc.) to traverse streams anywhere a road or a trail crosses a river. Some of our bridges are located exactly at these locations. Others (particularly those on the major river) are off main roads but show where foot traffic traverses the river (Figure 3).

3. We have adopted our own numbering system to combine together villages digitized from different sources. This numbering system does not match that of the MENRIS/ICIMOD data set.

4. This exercise will not consider planning units, such as administrative districts or planning blocks, although in an actual planning study, this information would also affect decisions of distribution.



Figure 3: Off-Road Foot Bridges in Dhading

- g) To examine the data files together, use DISPLAY Launcher to view DHADDEM with the palette file named NDVI. Then, with the Add Layer option in Composer, overlay RIVERS with the Uniform White symbol file. Use Add Layer again to overlay the second vector file ROADS with the Uniform Black symbol file. Finally, use Add Layer once more to overlay a third vector file SETTLMNTS with the Uniform White symbol file.

3. *Where are most of the settlements and roads found with regard to different components of the natural environment (hills, river valleys)?*

Now display LANDUSE with a Qualitative palette and legend and overlay again the same vector files as above (roads, rivers, and settlements) with the same user-defined symbol files.

Cost Distance Analysis

As discussed earlier, our task is to conduct a cost distance analysis of the area incorporating spatial characteristics of both the natural and social environments. The purpose is to estimate the accessibility of health services for village dwellers based on human costs of traveling by foot through the mountain terrain to the nearest health center. We will develop a model of the area using the available data. The model should tell us to which centers villagers would go, assuming that people choose least cost distance paths to the nearest health center and back to their villages.

We will calculate the average cost distance between each village and each health center. We cannot limit our analysis to one-way calculations of cost distance because in mountain environments effort spent varies greatly if one travels downslope instead of upslope and vice versa. After averaging the cost distance values for travel by foot in two directions, we can select minimum distance paths from every village to a health center.

Modeling this process requires several steps and the input of the disparate data listed above. Obviously, any computer model is just an approximation to the processes occurring in the real world. The algorithms used and variables included in the analysis determine the nature of the results. Some important factors which influence choices about traveling are difficult to model. The location of family and friends, fear of one doctor over another, dual purpose of journey (for journeys in mountains may be made rarely with a single objective) are not included here for lack of data and difficulty of incorporation.

The variables used, the availability, choice, and format of data, the method of presentation in quantitative form, as well as

the spatial analysis capabilities of available software will all directly affect the outcome. Their influence is no less important than that of factors related to accuracy and error in digital data and its sources. It is necessary to keep this in mind when interpreting the results and making policy recommendations. In our case, we will limit ourselves to the analysis of the data described above.

Preparing Input Files

Isotropic and Anisotropic Friction Images

As discussed earlier, the direction of frictional effects relative to movement is important in mountain environments. Here, the direction of slopes strongly affects the efforts needed to cross the area depending upon what direction one is moving -- up, down, or along a slope. Therefore, areas with the same *magnitude* of slope but with different *directions* of slope (aspect) might have very different frictions for movement in a given direction. VARCOST uses slope friction and direction values derived from raw slope and aspect values to calculate the friction of movement relative to direction. These data are combined to create an anisotropic cost surface.

Isotropic costs may be incorporated as well. VARCOST takes these into account by processing an additional friction surface which is similar to the ones used in the IDRISI modules COSTPUSH or COSTGROW. This surface shows features with friction values which are constant regardless of direction of travel: movement over roads, across rivers, and over different types of land cover. VARCOST processes both surfaces and adjusts anisotropic costs in each pixel accounting for isotropic frictions as well.

The resulting complex cost distance surface can be used to calculate pathways to and from specific locations. The route between the source area of the surface and any target locations can be derived. Relative quantitative figures can be computed for any source and target locations. These may be compared to decide upon optimal routes and locations.

In this case, we first create a source image of villages or health care centers. We then create friction surfaces for VARCOST. The first will be the slope magnitude/direction surface for anisotropic cost distance analysis. The second will be a combination of friction values for rivers, roads, bridges, and landuse types for the isotropic cost distance analysis. VARCOST executes the analysis of all the input data simultaneously.

Table 1: Costs of travel by foot

anisotropic	isotropic
slope magnitude	rivers
slope aspect	roads
	bridges
	landuse

Source and Target Images

For the purpose of the exercise, we limit ourselves to the analysis of availability of health posts or clinics (categories 3 and 4 in our typology) to villages. These facilities rank second in terms of services provided to the health clinic located in Dhadingbesi, and they provide services unavailable in the smaller health centers of lower rank. Therefore, we divide all settlements into two groups, those with health posts and those without them. This creates two sets of source and target pixels.

Based on these considerations, the first source image includes villages without health posts or health clinics. In order to

create it, we rasterize village point data, then we use the values file containing health center types to select the villages in the raster file with health posts.

- h) Run the module RASTERVECTOR from the Reformat menu. Specify STTLMNTS as the vector point file, STTLMNTS as the image to be updated, and use the default operation type (change cells to record identifiers of points). Click Yes when prompted to bring up INITIAL. Select DHADDEM as the image to copy parameters from and click OK.
- i) Now we have a raster image that contains unique village identifiers. Run the module ASSIGN with the feature definition image STTLMNTS, the attribute values file HEALTH, and the output image HLTHVIL. Look at HLTHVIL with the Qualitative palette.

This image shows the distribution of health facilities. The Edit/ASSIGN procedure renumbered the villages to have their health facility values rather than their village sequential identifiers. Different types of health facilities are shown with different color codes.

- j) Next, we want to reclassify these points so that the new file contains only those that belong to health categories 1 and 2 (those *without* health post facilities). Run RECLASS on HLTHVIL to create a new image called VILLAGES. Assign a new value of 0 to values *including* categories 3 and 4 (from 3 to just less than 5) and a new value of 1 to values *including* categories 1 and 2 (from 1 to just less than 3).

Currently, the villages without health centers all have a value of 1. In order to apply computer analyses to the villages individually, a unique identifier for each village is essential.

- k) To obtain unique identifiers for these points, we will use the module GROUP. Run GROUP from the GIS Analysis/Context Operators menu. Specify the input image as VILLAGES and call the output image VILLID. Click OK. This creates an image in which each village pixel is given a different arbitrary number (its own group). This image is the first source image that we use to calculate costs of walking from each village.
- l) We also need a vector version of this file for viewing purposes later. To create such a file, run RASTERVECTOR and specify the input image as VILLID and the output vector file also as VILLID. Use DISPLAY Launcher to view VILLID using the standard default symbol file.

Second, we need a target image representing the villages to which people will walk to obtain services of a health post or a health clinic.

- m) Run RECLASS on HLTHVIL and call the result CENTERS. Assign 0 to values including categories 1 and 2, and 1 to values including categories 3 and 4.
- n) As before, run GROUP to give each village coded 1 a unique identifier. Specify CENTERS as the input image and call the output CENTID. CENTID is the target image, containing villages with health posts and the health clinic. This image will also serve as the second source image, used to calculate costs of trips back to the villages from the health centers.
- o) Run RASTERVECTOR on CENTID and output a vector file of the same name. We will use it for visual examination later as well.

Preparing Friction Surfaces for VARCOST I: Anisotropic Costs

Magnitude of Friction Due to Slope

We next develop the friction surfaces. Slope angles are an important component of the model. We need to produce a slope image to use as a base for creating the anisotropic cost surface. We will use DHADDEM to create a slope image.

- p) Run the module SURFACE. Accept the default to calculate slope in degrees. Specify the input image as DHAD-DEM and call the slope image SLPDEG. Use DISPLAY Launcher to view SLPDEG with the default Quantitative palette, then use Cursor Inquiry Mode to explore the image. After examining it, you can open the image's metadata to view the minimum and maximum values of the image.

We include slope values in the cost distance analysis by assigning relative friction values to different slope angles. Ideally, these coefficients should reflect the actual human energy required to cross slopes in each pixel. On the other hand, as relative values, the friction coefficients need to make slope-related costs comparable to other kinds of costs involved.

There are a number of ways of assigning friction coefficients to slopes. The first is to reclassify continuous slope values into slope classes and assign a set of friction coefficients to different classes of slopes according to our assumptions. These would be based on experience about the amount of effort needed to cross them. This way is fast and simple, but the friction values are less consistent and quite subjective. The other method is to find an empirical relationship between steepness of slopes and human effort needed to overcome them. Once a relationship is derived and expressed in an equation, it is possible to turn the slope image into a continuous friction surface using mathematical transformations. If the dependency relationship is good, it is possible to argue that the resulting friction values are more consistent. This method also preserves the variation of the data because the values remain continuous rather than being broken into discrete and arbitrary classes.

In this case, we have only a small empirical data set from which to work. We will derive frictions ranging from 1 to 1000 from relative speeds of measured movement at different slopes. According to the results of a field experiment⁵, the average walking speed in the Middle Mountain area of Nepal is about 384 meters per hour. Regular walking speed on a flat surface is about 5,000 meters per hour. Slopes in the Middle Mountain region average around 20°.⁶ Flat surfaces have a slope of 0°. From this we may estimate that the friction of travel at 20° slope is around 13 times higher than travel at 0° slope. To begin with then, we will give 0° slope a friction value of 1, and 20° slope a friction value of around 13. We also know that the relationship between slope and movement is not a linear one. Observation shows that slopes approaching 40° and 50° become a prohibitive hindrance to movement, approaching impossible. Based on this information, a graph of the relationship between slope and friction should take a polynomial form (Figure 4).

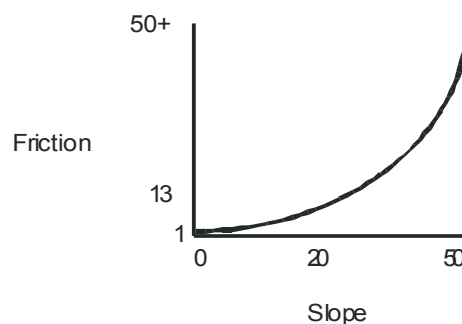


Figure 4: The Relationship of Slope to Friction

Based on this information, we fit a curve to the above points. In this case, we decided upon a simple binomial equation and using the above values derived:

5. From personal communication with Dr. Margaret G. Schmidt, Resource Management Sciences, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada.

6. This can be verified by using HISTO to look at the range of slopes between 1 and 40 in the DEM. The relationship is normal and centers around 20 degrees.

$$Y = .031X^2 - .025X + 1$$

Where Y is friction and X is slope. This means that at a slope of 0°, the friction is 1; at 20° slope, friction is about 11; at 40° slope, the friction is a prohibitive 42.6. As slope increases past 45°, friction increases exponentially. Obviously, this solution relies on a limited range of empirical data but does underline the way real world measurements of cost distance (in the form of relative speeds) can be incorporated to model friction magnitudes for slope in GIS.

Now the slope image must be manipulated to transform it into a friction image.

- q) To conduct map algebra in IDRISI, we use the modules SCALAR and TRANSFORM, both of which are found under the GIS Analysis/Mathematical Operators menu. SCALAR applies simple mathematical (i.e., multiplication, addition, division by a constant) operations uniformly to the values of all the cells of an input image. TRANSFORM conducts more complex operations (logarithms, roots, etc.) in a similar fashion. Run SCALAR and specify the input file as SLPDEG and the output file as TEMP. Select the multiplication option and enter - 0.025 as the scalar value.

To complete the next stage in the equation, run SCALAR again to add 1 to TEMP and name the result TEMP2.

Our next step is to take the square of X. To do this, run TRANSFORM and select the square option to take the square of SLPDEG and produce SLPSQ. Next, run SCALAR to multiply SLPSQ by 0.031 to produce SLPSQ2. Finally, run OVERLAY to add SLPSQ2 to TEMP2. Call the output image (the slope friction surface) SLPFR. Explore SLPFR with DISPLAY Launcher using the default Quantitative palette.

4. *What is the value range? How are these values distributed over space?*

Obviously, this way of calculating friction values is not the only possible solution. It is important to note that the choice of a technique should depend upon the method of travel (by foot or by vehicle, bicycle, train, etc.). Perhaps the age of the person traveling or the weight being carried might be incorporated for relating slope magnitudes to a friction surface. Whatever values are used in the slope friction image, they should be consistent with the range of friction values assigned to other spatial features (roads, land use, etc.). Slope will be combined with these other factors to determine cost distance so the compatibility of the coefficients is crucial.

Direction of Friction due to Slope

As discussed earlier, the direction of movement is another important issue. We model it using aspect -- the direction that a slope faces in each cell. A slope's friction value changes depending upon the direction a pixel is crossed relative to its aspect. The exact manner in which VARCOST alters the interpretation of friction values is determined by a function or set values used in the algorithm. The default in VARCOST is a cosine function designed such that it will transform friction into a force while moving downhill or assign a zero friction when moving laterally across a slope. This default may be altered using a user-defined function or set of values.⁷ These might be used to more carefully model movement, raising the friction of very steep downhill movement, for example. For simplicity, we use the default function here.

- r) To create an image of aspect values for the slope friction surface, run SURFACE and choose to calculate the aspect (output units will default to degrees). Specify the elevation model as DHADDEM and call the result ASPECT. Display it using the default palette and check values in the image with Cursor Inquiry Mode. Every pixel of an aspect image records aspect as an azimuth. This is the direction a slope faces in degrees from North. Most of the slopes facing similar directions are shown with similar colors. However, the Northern slopes with values around 0 and 360 degrees are indicated with the colors from both ends of the palette. Absolutely flat surfaces (where there is no slope and so, no aspect) are marked by -1 values. One more transformation is necessary before the aspect values are ready for VARCOST. VARCOST expects the friction direction values to be the

7. See the IDRISI Manual and the Anisotropic Cost Analysis chapter for details on creating user-defined functions.

direction of movement that encounters the *greatest* friction: the direction of highest resistance. In our case, the greatest friction is experienced when movement is straight up a slope. Aspect however, describes the direction which a slope is facing. This is the direction of downhill or *lowest* friction: the direction of least resistance. This is the opposite of what we want our direction image to show. Aspect must be reversed. To do so, we have to rotate ASPECT by 180 degrees. For example, where an east facing slope currently has an aspect value of 270, showing due West (the downhill direction) we want it to have a value of 90, showing due East (the uphill direction) (Figure 5). We do this by simply creating a temporary image, FLIP, and adding it to ASPECT. The possible range of values in both ASPECT and the output image will be from 0 to 360, but they should be 180 degrees different. We need to subtract 180 from all pixels with aspects between 180 and 360, and add 180 to all pixels with aspects from 0 to 180.

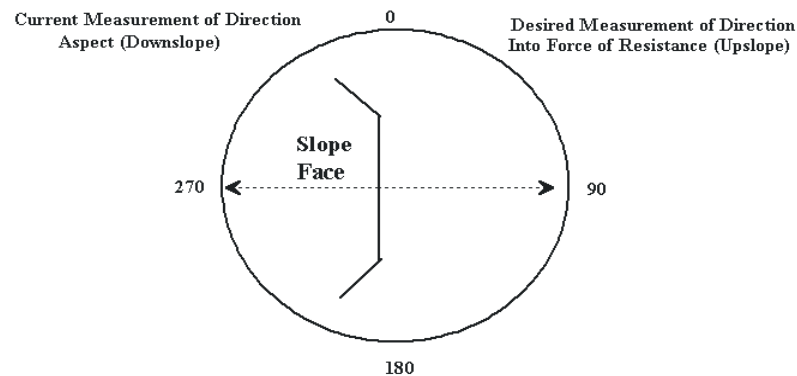


Figure 5: Turning Aspect 180 Degrees

s) Run RECLASS to reclassify ASPECT into FLIP as follows:

Assign a value of 180 to all values from 0 up to 180.

Assign a value of -180 to all values from 180 up to 360.

Assign a value of 0 to all values from -1 up to 0.

Next, run OVERLAY to add FLIP to ASPECT and produce the output image DIRECT. DIRECT contains values opposite to those in ASPECT.

Now we have both images needed for anisotropic cost analysis. These are the friction *magnitude* image, SLPFR, and an aspect-based friction *direction* image, DIRECT. We now require images showing the isotropic costs related to movement through various landuses.

Preparing Friction Surfaces for VARCOST II: Isotropic Costs

Recall that isotropic frictions are constant no matter what the direction of movement. In this analysis isotropic frictions are due to rivers, roads, landuse, and slopes over 40 degrees. Here, a separate friction image is created for each of these, in which friction values reflect the difficulty posed to a person walking through each feature. These are then combined into a composite image showing all friction for movement which is experienced isotropically.

Road Friction Coefficients

Let us create first a friction image that would reflect the quality of roads in a dry season, when the accessibility in the area is the highest. In this case, the differences in the size and quality of roads will determine the major differences in friction values. First, we need to rasterize the vector file ROADS.

- t) Run the module RASTERVECTOR. Specify the vector line file as ROADS and the image file to be updated as ROADS. Choose to copy spatial parameters from the image DHADDEM. Click OK. Display ROADS with a Qualitative palette.

Next we assign frictions to the image to make it a friction surface. The image shows that there are five types of roads. A range of relative friction values is needed to accommodate differences in their quality. We want these values to be lowest compared to other friction values in order to represent a tendency to stay on roads. We assume people are likely to walk along roads wherever it is possible.

Because the base friction value is equal to 1, this value is assigned to the highway. It is the easiest road to travel by foot. The dirt road is widely used for foot trips, although it is possible to drive a car on it as well. It receives the friction value 1.5. We assign a value of 2 to the major trail which is suitable for walking only. The remaining two kinds of trails are much more difficult to travel along, and we assign values of 3 to the larger trails and 3.5 to the smaller ones.

- u) Run Edit and enter the following numbers into the file.

```
1 1
2 1.5
3 2
4 3
5 3.5
```

When done, save as an attribute values file named ROAD. Select the data type for the file to be real (VARCOST requires that friction values be real numbers). Now run ASSIGN. Specify the feature definition image as ROADS, the values file as ROAD, and the output image as ROADFR.

The road network will be overlaid onto other landuses using a covering operation (as opposed to adding) to reflect the low costs of traveling along roads no matter what the other surrounding landuses might be.

Rivers as Barriers

Rivers of different sizes are an important part of the landscape of the Dhading area. Generally, river transportation is not very developed in mountains and people usually walk along river valleys. The greater suitability of valleys for walking should be taken into account by the slope friction image because slope values are less next to river beds, on adjacent terraces, and flood plains. Large and medium size rivers themselves, however, represent important spatial barriers unless there is a bridge or some kind of transportation. We assume that small rivers are not as inaccessible as large ones, and that people usually find ways of crossing them on foot. The available data on the location of bridges is not complete and we will also assume that a bridge or shallow water exists anywhere a road or trail crosses a river.

- v) Display the vector file RIVERS in DISPLAY Launcher with the Qualitative symbol file. Note that rivers of different sizes have different identifiers. Run RASTERVECTOR and specify the vector line file as RIVERS and the image file to be updated as RIVERS. Choose to copy spatial parameters from the image DHADDEM. Click OK. Then display ROADS with the Qualitative palette. All rivers, regardless of their size, are represented as lines of raster cells one pixel wide, but of different colors.

Before assigning friction values to rivers, we need to consider some issues in relation to the algorithm used in VARCOST

for calculating costs. First, as opposed to COSTGROW, VARCOST does not allow for the setting of absolute barriers, but only very large relative barriers. Second, as VARCOST determines distances from a source to each pixel, it calculates least cost paths by choosing the cell with the least value among eight surrounding pixels. Therefore, it may move in any of eight directions, including diagonally.

This presents a problem in the representation of rivers as barriers. The chain of pixels representing a river will form a barrier only where it is straight. Where there are diagonal shifts in the river, however, this river is crossed easily as the algorithm calculates least cost paths (Figure 6). This is not a problem for small rivers (we assume people can traverse them in many places), but the large rivers must be uncrossable. One way to mediate the problem is to widen the raster lines representing the large rivers to consist of three pixels instead of one. This will eliminate the possibility of diagonal gaps.

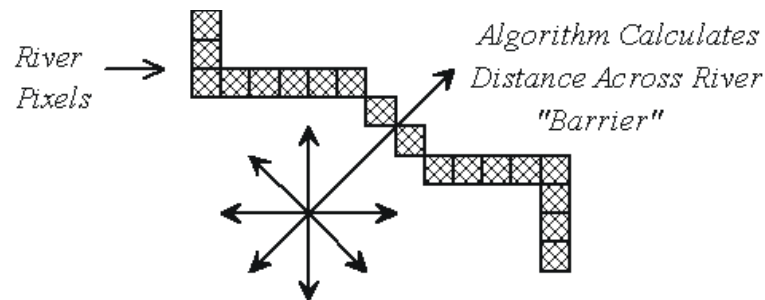


Figure 6: Calculating Distance Across River Diagonals

To thicken the lines, we run DISTANCE on an image showing only the large rivers. Then, based on the cell resolution, we reclassify the result to create a pixel wide buffer zone around each river. Afterwards, this image can be combined with that of the smaller rivers and friction values can be assigned to all of them. Note that in this case, we will sacrifice the accuracy of location and river width in order to increase the overall power of the model.

- w) Run RECLASS to create an image named RIVLARGE from RIVERS, assigning large rivers 2, 3, and 4 a value of 1 and the rest a value of 0. Run DISTANCE on RIVLARGE to create RIVDIST. Explore RIVDIST with DISPLAY Launcher using the default palette.

To create a river buffer that is one pixel wide around the original river line, we reclassify the distance data in order to set all distance values within the buffer to 1 and all those distance values outside the buffer to 0.

5. *How wide should the buffer zone be?*

- x) Run RECLASS on RIVDIST to make RIVBUFF. Reclassify values into a buffer zone where cells less than 120 m receive a new value of 1 and those 120 m and above receive a value of 0. Use DISPLAY Launcher to view RIVBUFF with the Qualitative palette. Verify that a three cell wide river was created. Make sure that no empty cells are found inside the river lines.

Next we must combine RIVBUFF with the other rivers. An easy way to do this is to use the Cover option in OVERLAY. Cover assigns the output image all of the values of the covering image except where the values are zero. In this case, the underlying values "show through" to the output image. We will cover the new wide rivers over the older ones while retaining all the other original river features. In OVERLAY, specify the first image as RIVBUFF, the second image as RIVERS, and call the result RIVERS1. Choose the Cover option (First covers Second except where zero). Look at RIVERS1 in DISPLAY Launcher with the Qualitative palette. All rivers now have the same value of 1, but the lines representing larger rivers consist of three pixels instead of one.

Finally, we are ready to convert this image into a friction image. Our task is to set rivers as barriers, so they will receive a very high friction value of 10,000. Relative to other spatial features, this means that it is 10,000 times harder to cross a river when no bridges are available than to move along a major road or on a flat surface. Small rivers will be given the same value as large ones. Traversability of these smaller streams is modeled through the presence of many diagonal "gaps" through which distance will be calculated.

- y) Open Edit and type in the following numbers separated by a space: 1 10000. Save as a values file named RIVERS1 with integer data type. Then run ASSIGN to assign the values file RIVERS1 to the image RIVERS1 and call the result RIVFR. RIVFR is a friction surface ready for use.

Bridges

We expect bridges or shallow water anywhere a road crosses a river. Bridges will have the same friction value as roads. At the same time, some bridges in the vector file BRIDGES are located on rivers far away from roads. We make these bridges part of the friction surface as well.

- z) Bring RIVFR on the screen with DISPLAY Launcher using the default palette. Use Add Layer in Composer to overlay ROADS with the user-defined symbol file ROADS, on top of RIVFR. Next, use Add Layer again to overlay a second vector file BRIDGES, this time with the standard default palette symbol file. As mentioned before, some of these bridges are far away from roads. Determine which bridges these are by windowing out small portions of the image.

Visual examination makes another problem clear. We have thickened the large rivers to enhance their importance as spatial barriers. But since the locations of bridges are represented by points, all of them now "hang" in the middle of the now widened rivers. Therefore, if we rasterize bridges as points and assign friction coefficients, they will fall into the middle of rivers and be surrounded by pixels with very high friction. They would not play any role in the final friction surface. One possible way of dealing with this issue is to "extend" the bridges to the shores by transforming them from points into lines so that they will link the opposite river banks.

We will do this using the onscreen digitizing option. The three toolbar icons you will use for onscreen digitizing are the Digitize icon (crosshatch), the Delete Digitized Feature icon (X), and the Save Digitized Data icon (downward arrow). Each of these three icons are located together in the middle of the toolbar. Be sure you know where each of these icons are before proceeding.

- aa) With the RIVFR image still displayed with the vector files ROADS and BRIDGES, window out a small portion of the image containing one bridge using Zoom Window. Decide how to extend the bridge. We suggest that the lines representing bridges go in vertical or horizontal directions as much as possible, because VARCOST gives these directions higher priority.

Press the Digitize icon to activate the onscreen digitizing mode. Name the file you will create BRFRIC, choose line as the type of vector feature to digitize, and use the default feature ID of 1. Click OK. The pointer for the mouse will now be replaced by the Digitize icon. Digitize the new line for the first bridge by first moving the cursor to the starting location on the edge of a river. Press the left mouse button to start, then move the mouse to the opposite side of the river, and press the left mouse button again (Figure 6). A dashed line should appear. Then press the right mouse button to finish the line (it will turn solid). If the digitized line is incorrect, reject saving the feature by pressing the Delete Digitized Feature icon.

Continue by windowing to another portion of the river containing a bridge. When you select the Digitize icon a second time, choose the default option of adding features to the currently active vector layer. Click OK. Change the number in the ID or Value box to 1. Don't worry if you make a mistake--just delete the feature and digitize the same bridge again. After all eight bridges are done, press the Save Digitized Data icon and click Yes on the dialogue box that appears.

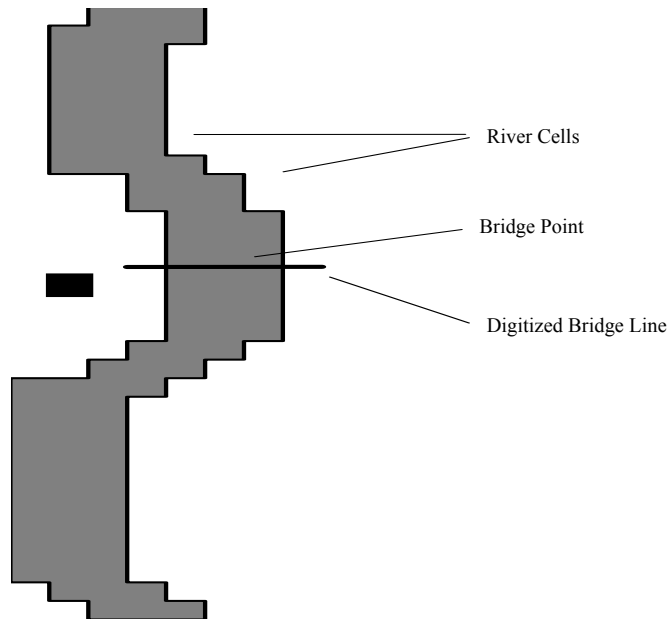


Figure 7: Digitizing Bridges

- ab) Look at the vector file BRFRIC in DISPLAY Launcher to verify the information you have created. You may want to display the file BRIDGES next to BRFRIC to verify that you have digitized a line for each bridge by dragging one of the images to the right. Run the module RASTERVECTOR. Specify the vector line file as BRFRIC and the image file to be updated as BRFRIC. Choose to copy spatial parameters from the image DHADDEM. Click OK. Display BRFRIC with the Qualitative palette.

This image is ready to be converted to a friction surface. We want to assign a friction coefficient of 1 to bridges to stress their importance in ease of use and priority to local travel. Since the values of the bridges are already equal to 1, all that is necessary to make this a frictional value is to convert it from the integer data type to the real number data type (as required by VARCOST).

- ac) Run CONVERT from the Reformat menu on BRFRIC. Choose the real binary format for the output file, the default output name BRFRIC, and all other defaults. When prompted that the BRFRIC file already exists and that it will be overwritten, click Yes.

Land Cover

Obviously, land covers vary in terms of how easily they are crossed. Among the landuse types recorded in LANDUSE, forested areas are more difficult to traverse than open agricultural lands in valleys or pastures. We need to assign friction values to them according to these differences. One possible way is the following:

Table 2
Landuse Types and Friction Coefficients

Landuse Type	Landuse Code	Friction Coefficient
Hillslope agriculture	1	15
Valley agriculture	2	7
Grazing land	3	5
Non-agric. land	4	5
Forest	5	30

This scale tells us, for instance, that it takes 30 times greater effort to go through a forest, 15 times more effort to cross a hillslope agricultural field, and just five times larger effort to walk through a grazing field than to travel on a major road. The difference between forested areas and grazing land is also large.

ad) Use Edit and ASSIGN to assign the above values to LANDUSE and create a new image LANDFR.

Combining Friction Surfaces

All components of the final friction surface are ready. It is now necessary to combine those components that will constitute the isotropic friction surface. The summary of what was generated is in the table below:

Table 3
Images for Isotropic Cost Analysis

Image Name	Range of Values
ROADFR	1-3.5
RIVFR	10000
BRFRIC	1
LANDFR	5-30

A combination of adding and covering operations in OVERLAY integrates the images designated to isotropic cost analysis into one final friction surface. The sequence of overlays is very important, because it is necessary to make sure that no information is lost. For instance, while friction values in LANDFR and RIVFR are additive, roads and bridges must be overlaid last, using a cover function.

ae) You might want to look at the intermediate images in DISPLAY Launcher (be sure to use a qualitative palette) to gain a sense of how the friction surface is created. First run OVERLAY to add LANDFR and RIVFR to produce SUR1 (First + Second). Then run OVERLAY to cover ROADFR on SUR1 to produce SUR2 (First covers Second except where zero). Finally, run OVERLAY again to cover BRFRIC on SUR2 and create the resulting isotropic friction surface SURTOT.

We now have a friction image incorporating all of the land isotropic features which slow or hasten movement through this environment. Coupled with the anisotropic friction surface and source locations, we can now calculate cost distances across this diverse landscape.

Running VARCOST

The following images are required by VARCOST:

- 1) a source image (containing the features from which cost distances are calculated);
- 2) a magnitude image for the anisotropic friction surface;
- 3) a direction image associated with the magnitude image;
- 4) an isotropic friction image.

We have all of them. CENTID and VILLID are our source images, SLPFR is the magnitude image, DIRECT is the force direction image, and SURTOT is our composite isotropic cost surface.

The result of VARCOST will be a cost distance surface which shows the accumulation of costs traveling from source points to the other villages. Every pixel of the resulting image will store a value equal to the minimum cost of moving to this cell from the nearest source point. The density of both source and target features, though, affects the outcome of cost distance values. To illustrate the problem, let us use VARCOST with CENTID as a source image. This way, we will attempt to calculate the cost distance surface from all the health centers at one time. Then we will compare the result to the distribution of villages and judge its efficiency.

- af) Run VARCOST from the GIS Analysis/Distance Operators menu. Specify CENTID as the source image, SLPFR as the friction image, and DIRECT as the direction image. Click on Use isotropic friction surface and specify SURTOT as the image. Call the resulting output image COSTTRY. Use the default Cosine function and enter an exponent of 2. Click OK. Look at COSTTRY in DISPLAY Launcher using the default palette.

This is an image which shows the cost distance away from all of the centers. Each cell has the value of the cost distance away from the *nearest* health center. One possible use of this image might be to assign cells to their closest center (in terms of cost distance). This might provide some sense of which villages fall within the closest walking range of which centers.

- ag) Run the module ALLOCATE from the GIS Analysis/Distance Operators menu. Specify the distance image as COSTTRY, the target image as CENTID, and call the result COSTRGN. This will assign each pixel in CENTID to the closest center based upon cost distance in COSTTRY. Look at COSTRGN in DISPLAY Launcher using the Qualitative palette. Using the Add Layer option in Composer, overlay the vector file VILLID on COSTRGN with the default symbol file.

Each pixel in this image has been given the value of the identifier of the center to which it is closest. As we can see, it has created cost distance "catchment" areas around health centers. Villages can be seen falling into the walking "distance-shed" of the various centers. To see which particular villages fall into the closest range of these centers, we may extract the values of the pixels in which the villages are located. In this way, we may "assign" villages to their nearest health centers. It is important to keep in mind this distance is measured *from* the health centers to the villages. The reverse direction may differ significantly as we will see.

- ah) To do this, run the module EXTRACT using VILLID as the feature definition image and COSTRGN as the image to be processed. Select minimum as the summary type. Choose attribute values file as the output type and name it VILCENT. Examine the VILCENT values file with Edit.

This file shows the village identifier on the left, with its nearest center (measured in distance *from* the centers) shown on the right. Notice that many centers have been allocated multiple villages.

6. *How many villages fall within the least cost distance range of center #2? Center #9?*

From this analysis, we can see where some centers may be working past their capacity, while others serve relatively few villages. This analysis could be further expanded by totaling the populations of the villages within the range of each health center and comparing the demands placed on each center.

Leaving this avenue of analysis aside, our next step might be to determine how far these villages are from their nearest centers.

- ai) To do this, run EXTRACT using VILLID as the feature definition image, and COSTTRY as the image to be processed. Use the minimum summary type. Again, choose a values file as the output type and name the file VILDIST. Examine the VILDIST values file using Edit.

This file shows the village identifier on the left, with the distance from the nearest center (measured in distance *from* the centers) shown on the right. Some villages are comparatively close to health centers (as in the case of villages 45 and 58), while others are quite isolated (like villages 54 and 71).

To further explore the walking route between centers and villages, we may derive the least cost pathway between a given village and its nearest center. In this case, we will look at village 50. Its closest neighbor, center #12, and relative distance are found in the values files VILCENT and VILDIST created before. To find its least cost pathway, we must first isolate it into its own raster image, and then calculate the path.

- aj) To do this, use Edit to create a values file named VILLID that will assign village 50 a value of 1. Use ASSIGN to assign this file to the image VILLID and create a new image called VILL50. In this image, village 50 will receive a value of 1 while the other villages will default to 0 and disappear.
- ak) Next run the module PATHWAY from the GIS Analysis/Distance Operators menu and specify COSTTRY as the cost surface image, VILL50 as the target image, and call the output file VIL50PTH. Look at VIL50PTH using DISPLAY Launcher and the Qualitative 16 palette.

This is the least cost path *from* center 12 to village 50. The relative distance along this path is described in VILDIST. The return distance, from the village *to* the center, may be considerably different. Calculating the cost distance in the other direction requires use of the VARCOST module again. In this case, we will run the module using this village as the only source pixel, deriving all distances from this location.

- al) Run VARCOST and specify VILL50 as the source image, SLPFR as the friction image, DIRECT as the direction, and SURTOT as the isotropic friction surface. Use the default Cosine function and enter an exponent of 2. Call the output COST50 and click OK. Use DISPLAY Launcher to view COST50 with the default Quantitative palette.

This is an image where each pixel is given a value based on its cost distance from village 50. To determine the nearest health center, we may extract the distance values for all the centers.

- am) Run EXTRACT using CENTID as the feature definition image and COST50 as the image to be processed. Choose a values file as the output type and name the file DIST50. Select minimum as the summary type. Examine the values file with Edit.

7. *Which center is closest to village 50 based on walking from the village to the center? Is this the same as the previously identified nearest center going the other direction?*

This underlines the anisotropic character of movement in mountain environments. The distance of the trip out is different from that of the walk back. We conclude that we need to take into account both directions in order to determine which health post is actually closest to a given village. In this case, we might assign a village to another health center, the one for which the *average* cost distance values are lower. To find out the average cost distance between these villages and

health centers, it would be necessary to extract distances between all villages and all health posts in both directions and average the results.

Conclusion

Similar queries are now possible. Any combinations of locations can be examined in terms of travel in two directions. Pathways can be drawn and marginal locations (those having relatively high two-way travel distances) can be identified and targeted for support. The inputs of the model can also be varied to account for different methods of transportation and different sources of friction.

These analytically powerful two-point queries suggest other, more complex possibilities. Anisotropic cost distance allows a level of modeling which opens up larger optimization questions. In this case, it is a natural extension of the analysis to ask what changes in the cost distance surface might occur with the addition of a new health care post. Where is the optimal location for a new center? The answer is in no way immediately evident from the discrete sets of input data. It is a question ideally suited for computer modeling.

There are several approaches to seeking a solution to this question. These would involve the implementation of repeated cost distancing routines in GIS combined with statistical management of resulting data sets. The length and complexity of the routines involved does not allow its inclusion here.

Exercise 4: Modeling the Impact of Mountains on Regional Climate



Introduction

This exercise explores the relationship between topography and climate by modeling the distribution of solar energy insolation to predict the locations of high convection and precipitation. As a result, it is an apparently complex exercise, incorporating a number of climatological indices and working to derive them from the slopes and angles of a mountainous environment. The fundamental GIS principles behind the exercise are not, however, overly complex. The bulk of the analysis is simple map algebra which derives new layers of data (heat islands) by substituting raster images (slope, aspect, and latitude) for variables in climatological equations. The exercise underlines the extent to which basic GIS analysis can be harnessed to address highly complex problems.

The Region and the Problem

Solar energy drives several climatological, hydrological and biological processes on or near the earth's surface and is therefore an important input for modeling these phenomena. The estimation of spatially distributed actual and potential solar radiation facilitates a broad range of applications. In this exercise, we explore the problem of identifying heat islands over the mountains of the Puna Atacama area of southern Chile. These heat islands indicate locations for potential sources of moisture for the region. To locate these heat islands, we estimate the level of incoming solar radiation which is referred to as insolation potential. Since the insolation potential is affected by the slope of the earth's surface, we predict the intensity of radiation falling over a mountainous region by employing a digital elevation model.

The exercise forms part of a project undertaken by the Geography Department of the University of Chile working in collaboration with the University of Bern, Switzerland. The project aims at defining and explaining the sources and availability of water in the Puna Atacama province of Chile, in an attempt to find a solution to the region's scarcity of water problem.

Puna Atacama is an area in Chile located in the extreme arid zone of the Atacama desert. It is a region of high mountain relief with peaks ranging in height between 2250 and 6000 meters above sea level. In the past, the region supported ancient civilizations that adapted well to prevailing environmental conditions. In recent times, the harsh relief and extremely dry climate have made the region less attractive to human settlement. The native population earns a living from subsistence agriculture and animal husbandry.

The region is important for its mineral production, especially copper mining which provides the backbone for Chile's ailing economy. The mainly open cast mining activities taking place in the region require the use of great quantities of water during the production process. Mining activities also generate tremendous urban growth among the few urban centers, Antofagasta and Calama, the capital of the Loa province. The domestic consumption of water has therefore increased in recent times adding to the already acute situation of water scarcity in the region. It was estimated that by 1990, the population had exploited nearly 95% of the surface water. Already, the region's subterranean water, currently exploited in some mining centers, possesses either very little or no recharge capacity. It seems unlikely that the available water will be sufficient to supply farming and livestock breeding and continue to be used in the mines.

The extremely arid conditions of the Atacama region stand in great contrast to the heavy summer precipitation experienced on the slopes of mountains to the northeast of the region. The rainfall experienced in the northeast has its origins in the continental air circulation resulting from the Bolivian High Pressure belt. An intense summer radiation in the Atacama region creates heat islands over the mountains. This leads to the development of a strong vertical temperature gradient and a condition of intense low pressure which draws in cool and wet winds from the sea into the highlands region. Over the highlands, a combination of climatic conditions, including strong convection currents, makes possible the advection of moisture-laden air masses from the Atlantic ocean which develop into rain. Precipitation in the region is thus related to convective processes concentrated on heat islands that function as thermal accumulators to attract maritime air masses over mountains.

The Exercise

In this exercise, we will attempt to model surface temperature and atmospheric conditions of the Atacama region as a means of facilitating the location of heat islands that serve as indirect sources of moisture for the region. In order to model heat islands, we will need to define the Instantaneous Insolation Potential (IIP), a measure of the intensity of solar radiation at any single point at any moment. Insolation is determined by surface locations and conditions as well as atmospheric conditions. The model will require the following:

- 1) Topographic Information. Slope and latitude information will be entered into a formula to model insolation variation.
- 2) Atmospheric Turbidity and Density Information. The thickness and movement of the atmosphere in this location will be postulated and incorporated to model the proportion of the solar constant reaching the surface.

Using this information, we model the location of heat islands and examine which drainage nets they nourish in the region.

The following data are used in the exercise:

AREADEF	an outline map of the study area
CDEM	a digital terrain model derived from digitized contour lines
LATTITUD	a file showing the latitudes (parallels) of the region
THRDTM	a three dimensional perspective image of the study area
NOAAFEB	a February 11, 1986 thermal image of NOAA satellite
NUBE4	a Boolean image of clouds in the above satellite data
INSOP	a vector map of a section of the study area

The mountains of the study area are clearly visible on the exaggerated relief display of THRDTM.

- a) For an overview of the study area, use DISPLAY Launcher and the user-defined palette THRDTM without a legend to view THRDTM, a three dimensional image of the area. Notice the broad valley in the south where a large salt lake is located. To the northeast are very high mountains over which cool wet winds from the Atlantic ocean blow.

The exercise is divided into two sections. The first part reviews the mathematical models used, and the second demonstrates the implementation of those models in GIS.

Equations for Modeling Insolation Characteristics

The equations incorporated in this exercise are used to determine the intensity of solar radiation or insolation at any point and time. As a first step, they are introduced and discussed here. In the second part of the exercise, raster images will be used to substitute for variables in the equations to create surfaces showing the variation in insolation potential in the region.

The Insolation Potential Formula

These formulae are based on equations using principles of spherical trigonometry¹. The principal equation explaining insolation characteristics of an area is given below:

$$(I) \quad IIP = kp^e \cos v$$

where,

IIP = Instantaneous Insolation Potential, representing the quantity of insolation energy received for each surface unit measured in langley/min/cm

k = solar constant corresponding to approximately 1.98 langley/min for a plane lying perpendicular to the ray of the sun

p = atmospheric turbidity index

e = atmospheric density penetrated by the sun rays

v = angle of insolation for a planar surface in any position

Put simply, this equation states that the insolation in any place is the solar constant (a constant quantity of energy reaching the outside of the earth's atmosphere) modified by the turbidity (movement) and density (thickness) of the atmosphere, and the angle at which the radiation strikes the surface of the earth. We begin with the last of these.

Insolation Intensity and Zenith Angle

The angle of insolation is determined by the angle of the surface slope being struck by sunlight as well as the angle of the sun at the time of illumination. Significantly, as rays of sun fall in less than perpendicular angles to a planar surface, the energy received progressively decreases; the more oblique the illumination, the less intense the energy. In general mathematical terms, whatever the position of a planar surface, insolation intensity is the inverse of the angle formed by the sun-ray and the normal (perpendicular) to the plane (Figure 1). We will call this angle v . It is this angle we need to solve for, given slope and the sun's position at any time. The slope of the plane is here defined as α .

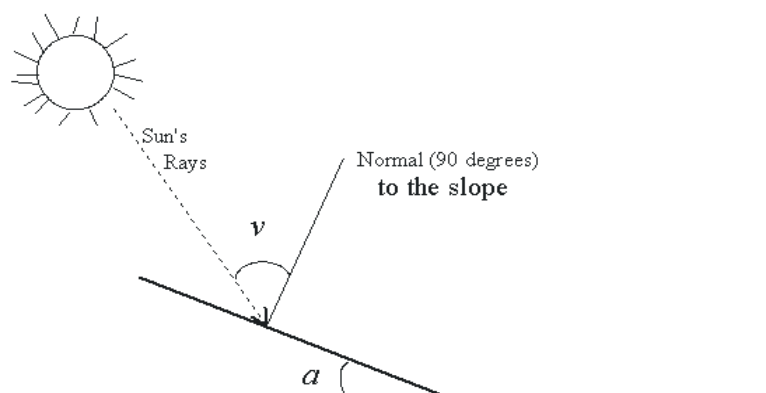


Figure 1: Vertical Parameters: Slope of a Plane and Angle of the Sun's Rays

1. For more information on the use and development of these general equations, see Barry, R.G. 1981. Mountain Weather and Climate. London: Methuen.

Clearly, the angle of the sun's rays (ν) is a function of α . It is also a function of the horizontal position of the sun in the sky relative to the direction in which the slope is facing. To solve for ν , we require the position of the sun in the sky as it casts its rays and a description of the plane's slope relative to that position. To describe both the slopes facing and the direction of illumination in a formula together, we use their azimuth coordinates. The azimuth coordinate is the point on a 360° arc where the plane or the sun's line of illumination intersects with the horizon. The difference between the two coordinates is described as d (Figure 2).

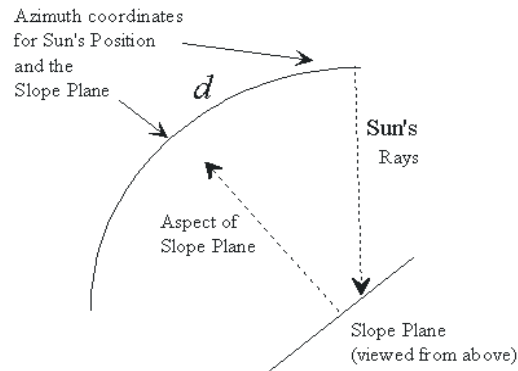


Figure 2: Horizontal Parameters: Difference in Azimuth of the Slope Plane and the Sun's Rays

To better understand this, imagine north facing slopes in the northern hemisphere. They receive less direct radiation than south facing ones. This would be represented by a high value for d .

Now, given equation I, and taking into consideration the fact that the zenith distance of the sun varies constantly, angle ν is derived from:

$$(II) \quad \cos \nu = \cos \theta \cos \alpha + \sin \theta \sin \alpha \cos d$$

where

θ = zenith angle of the sun

α = slope of the plane

d = difference between the azimuth of the plane and the azimuth of the sun

The zenith angle of the sun is a measurement of the height of the sun in the sky, described in degrees away from a point directly overhead. The angle varies from 0 degrees at midday to 90 degrees at sunrise or sunset. The zenith angle is usually determined during midday when the sun is at its height.

The zenith angle of the sun will of course be different at any day or latitude. An additional formula is required to derive θ . For any date or latitude, this angle is obtained by the formula:

$$(III) \quad \theta = (-1) L \pm D$$

where

θ = zenith angle

L = latitude of plane (negative in southern hemisphere and positive in the northern hemisphere)

D = solar declination (positive or negative)

The solar declination is the angular distance of the sun either north or south of the equator. This is a function of the time of year. It is independent of the location of the observer and can be found for any given date in a naval almanac or other source².

In summary, we require latitude (L) and solar declination (D) to calculate the sun's zenith angle (θ). With the zenith angle (θ), the slope of the surface (θ), and the difference between the sun's azimuth and the plane's azimuth (d), we can calculate ν for use in deriving insolation potential (IIP). We now proceed to the qualities of the atmosphere which influence insolation and which are incorporated into formula I as turbidity (p) and density (e).

Atmospheric Turbidity and Density

If we assume there is no atmospheric interference to incoming sun rays, then the energy received on a horizontal (non-sloping) surface will be simply defined by:

$$E' = k \cos \theta$$

where

k = solar constant

θ = zenith angle

In reality, however, even with the sun at its zenith, only an estimated 0.7 of the solar constant reaches the earth's surface. This is a result of atmospheric effects: turbidity and density. Turbidity is the convection of the earth's atmosphere and is inversely related to insolation transparency. Under conditions of high turbidity (corresponding to minimal transparency), only 0.3 of the incident energy reaches the upper limit of the atmosphere. Turbidity is never stable but oscillates between its maximum and minimum points.

Density is simply the thickness of the atmosphere. It is also inversely related to atmospheric transparency. The density is determined by a number of factors, including the location of the observation (latitude) and the zenith angle of the sun. Latitude accounts for variation in the thickness of the tropopause, a layer of the atmosphere which is much thicker at the poles than at the equator. The zenith angle of the sun determines the angle which the sun's rays will contact the uppermost limit of the atmosphere. This will in turn determine the thickness of the atmosphere through which the solar energy must pass. As the angle becomes more oblique, the sun's rays must pass through more atmosphere to reach the point of observation.

A number of complex formulae are available to model the density of the atmosphere at a given point. These include variables which account for the ratio of the geoid to the surface of the tropopause at any given latitude, as well as angular variables accounting for the angle at which the sun strikes the outer atmosphere. For a full discussion of these variables and models, refer to Gimpel (1987).

If we assume the lowest limit of the atmosphere (earth's surface) and the top of the atmosphere to be horizontal and parallel to each other (eliminating concern for variations in the tropopause), then for zenith angles under 60 degrees, value e can be simply determined from:

$$(IV) \quad e = 1 / \cos \theta$$

where

2. By making substitutions from equation III into equation II above, we could derive a formula to compute the zenith angle based on the latitude, the declination, and the angle hour of the sun away from observed noon. For further discussion see Barry, 1981: pp.68-69.

e = atmospheric density (thickness)

θ = zenith angle of the sun

In the case of zenith angles greater than 60 degrees, the parallels will progressively diverge from one another, and more complex solutions are required.

Finally then, the energy that falls onto a horizontal (non-sloping) surface after penetrating the atmosphere (E) can be defined by:

$$(V) \quad E = E' p^e$$

where,

E' = energy received by a horizontal surface without atmosphere

p = Turbidity Index of the atmosphere

e = density/thickness of the atmosphere

So by determining or estimating the turbidity of the atmosphere at the time of observation (p), and deriving the thickness of the atmosphere based upon the zenith angle of the sun (e), we can account for atmospheric effects on the insolation in a given area. By further incorporating slope effects as explained above, we introduce angle ν and its components: zenith, slope, solar declination, and latitude. Together, these are encompassed in formula I, as given at the beginning of the section.

Now that we have explored the derivation of the models, we will apply them to a real world situation using the data images listed at the beginning of the exercise.

Modeling Instantaneous Insolation Potential in GIS

The general formulae explained above can be used to determine the Instantaneous Insolation Potential at any point at any given time. By substituting raster images (of slope and latitude) for individual variables in the formulae, we can determine the pattern and variation in insolation across a region. Clustered "heat island" areas of high insolation should become readily apparent in this derived insolation image. These can be used to further focus the search for water resources in the region.

We will estimate Insolation Potential across the Atacama region for each pixel at midday time. The midday time is preferred because at that time, it is possible to obtain maximum insolation with very little zenith angle and with minimum atmospheric interference. To do this, we use the initial equation:

$$(I) \quad IIP = k p^e \cos \nu$$

which is derived from three major components -- atmospheric conditions, topographic characteristics, and the angle of the sun. We will return to atmospheric effects shortly. For now, we will concentrate on equations II and III to derive the insolation intensity for the region: $\cos \nu$. The major components required for computing $\cos \nu$ are the meridian's zenith angle, solar declination, the slope of the plane, and the cosine and secant of the angle the sun ray makes with the normal of the plane's surface. These determine ν through the relationship:

$$(II) \quad \cos \nu = \cos \theta \cos \alpha + \sin \theta \sin \alpha \cos d$$

$$(III) \quad \theta = (-1) L \pm D$$

Estimating Insolation Intensity ($\cos \nu$)

Latitude (L)

First, it is necessary to derive the zenith angle as in equation III. As a first step, we created an image to show the latitudes of the region. First, the dimensions of the study area are determined, and a pixel resolution of 1 km chosen.³ At this resolution, the region requires 210 columns and 360 rows.

- b) To view the result of this process, use DISPLAY Launcher to examine the file LATITUD and with Cursor Inquiry Mode, examine the region's values. All other images for the exercise will take on the dimensions and parameters of LATITUD.

Solar Declination (D)

In addition to the latitude, we also need the solar declination to determine the zenith angle. Due to the revolution of the earth around the sun, the height of the sun in the sky at noon varies throughout the year at any latitude. The rays of the sun are directly overhead at the tropics of Cancer (23.5N) and Capricorn (23.5S) during the summer and winter solstices of June 21 and January 21 respectively. Between these times, the point of direct radiation travels steadily North and South across the equator, moving daily. The location of this point on any day, in degrees north or south of the equator, is defined as the solar declination⁴.

The solar declination angle (for a location in the area) for the day of concern (February 11) is $-14^{\circ} 18' 21.9''$ (-14.30608063)⁵.

Zenith Angle (θ)

We now can derive the zenith angle of each pixel according to the steps in equation III by setting and using LATITUD for latitude values. By using SCALAR twice, latitude and the solar declination angle are transformed into the zenith angle.

- c) Run SCALAR, specify the input file as LATITUD and the output file as LAT2. Select the multiplication operation and specify the scalar value of -1. Click OK. Repeat the procedure with SCALAR by adding -14.30608063 to LAT2. Call the output image ZENANG. This image contains the meridian zenith angle per pixel for midday of February 11th.

Cos ν

Returning to equation II, we want to calculate the *sine* and *cosine* of the zenith angles. However, it is necessary to convert the zenith angles into radians before computing the sine and cosine. The TRANSFORM module handles all three steps.

- d) Run TRANSFORM and select the transformation type to be Radians (x in degrees). Specify the input file as ZENANG and call the output ZENRAD. Run TRANSFORM a second time, and select to compute the Sine of ZENRAD. Call the result ZENSIN. Note that the values will appear the same. Finally, run the module again,

3. This spatial resolution approximates that of the thermal sensor of the NOAA satellites. One of the original objectives of the study is to find the correlation between the estimate of the insolation intensity and an image of a thermal band of NOAA satellite data. The estimates of instantaneous insolation potential are comparable to thermal emissivity once the effect of different land use emissivities are removed.

4. Following the above formula, for areas between the tropics, on the days when the declination is the same as the degree of latitude at a given location, the sun passes directly overhead; the zenith is zero. This is the principle behind the traditional nautical estimation of latitude at sea.

5. The date of the original NOAA satellite observation.

choose to compute the Cosine of ZENRAD, and call the output image ZENCOS.

These fit into the equation here:

$$\cos \nu = \cos \theta \cos \alpha + \sin \theta \sin \alpha \cos d = ZENCOS \times \cos \alpha + ZENSIN \times \sin \alpha \cos$$

Azimuth Angle (d)

Among the variables defining the Insolation Intensity in equation II is d which represents the difference between the azimuth of the plane and that of the sun. First, we will need the azimuth position of the sun. Since we chose to evaluate the Insolation Intensity at noon for the midsummer day (February 11) in a region located in the southern hemisphere, the solar azimuth angle for all pixels will be zero. This is because the sun does not reach its height on this day but does pass horizontally due North (0° azimuth) through the sky at noon (a few vertical degrees from directly overhead). Therefore, the difference (d) between the azimuth of the sun (in this case 0) and the azimuth of the slope faces in the area will simply be equal to the azimuth of the slopes; their aspect.

- e) Use DISPLAY Launcher to view CDEM, a relief model of the study area derived from interpolated contours, with the default Quantitative palette. Use Metadata to examine the documentation file for CDEM. Notice the flag value and flag definition fields. In this case, 0 is a special value. Since it is designated as background, it represents any pixel beyond the extent of the study area. Any pixel with a value of 0 will be left out of slope and aspect calculations. This prevents interpolations beyond study area values which would produce extreme or irregular values at the edge of the study area.
- f) Run the module SURFACE and choose to calculate the aspect. Specify the input elevation model as CDEM and call the output aspect image EXPO.⁶ Retain the defaults for the other options and click OK. Next, run HISTO and specify the input file as EXPO and click OK. You will notice that some portions of the image contain negative values.

1. *Explain why negative values occur in the aspect image.*

We need to convert the azimuth angles contained in the aspect image into radians and then compute the cosine of the image as we did earlier with ZENANG.

- g) To do this, run TRANSFORM and select the transformation type to be Radians. Specify the input file as EXPO and call the output EXPORAD. Repeat the procedure and compute the Cosine of EXPORAD. Call the resulting image EXPOCOS.

This fits into the equation here:

$$\cos \nu = \cos \theta \cos \alpha + \sin \theta \sin \alpha \cos d = ZENCOS \times \cos \alpha + ZENSIN \times \sin \alpha \times EXPCOS$$

Slope (α)

The last variable in equation II is $\cos \alpha$ representing the slope of the surface plane.

6. Deriving an azimuth of a one kilometer width pixel assumes that the value generated approximates the pattern of azimuths from a larger pixel resolution covering the same one kilometer area. Testing and research of this assumption is required. The fractal dimensionality of the area as well as the extent to which a region gradually increases in height from one direction may affect this. Research incorporating autocorrelation figures for decreasing resolutions is one suggested area of research.

- h) To create a slope image, run the module SURFACE and specify that you want to determine the slope from CDEM. Call the resulting image DSLOP and click OK. Run TRANSFORM to convert the degrees of slope in DSLOP to radians. Call the result SLOPRAD. Repeat TRANSFORM to compute the Sine and Cosine of SLOPRAD and call the results SLOPSIN and SLOPCOS.

These fit into the equation here:

$$\cos \nu = ZENCOS \times SLOPCOS + ZENSIN \times SLOPSIN \times EXPOCOS$$

Having computed all the variables in equation II, we will now combine them to produce the potential insolation intensity. We will make use of a macro file to save time and effort in performing the required operations. Run Edit and open a new file. Type in the following lines in the file:

```
overlay x 3*zencos*slopcos*zenslo
overlay x 3*zensin*slopsin*slozen
overlay x 3*slozen*expocos*exposen
overlay x 7*zenslo*exposen*insopot
```

Save the file with the name COSV. In the “Save as type” box, select “Macro file (.iml)”. Each line in the macro represents the following operations:

- | | | | |
|---------------------|---|---------|-------------------------------|
| 1) ZENCOS x SLOPCOS | = | ZENSLO | ($\cos \theta \cos \alpha$) |
| 2) ZENSIN x SLOPSIN | = | SLOZEN | ($\sin \theta \sin \alpha$) |
| 3) SLOZEN x EXPOCOS | = | EXPOSEN | ($SLOZEN * \cos d$) |
| 4) ZENSLO + EXPOSEN | = | INSOPOT | ($\cos \nu$) |

This reduces Equation I to:

$$IIP = k p^e (INSOPOT)$$

- i) From the File menu, select Run Macro and specify the input file COSV. Next, use DISPLAY Launcher to view INSOPOT with the default palette.

Modeling Atmospheric Conditions in GIS

Now that we have completed estimating the insolation intensity under conditions of no atmosphere, we will model the atmospheric variables that affect the insolation in the region. We will then combine the two images to produce an image of the total potential insolation for the region. As was done in the previous section, we will apply the relevant equations to create images simulating the atmospheric density and the turbidity. Finally, atmospheric density and turbidity images will be combined with INSOPOT to produce the insolation potential.

Atmospheric Density (e)

We can derive atmospheric density, e , using a variety of formulae. For conditions in which the zenith angle is less than 60 degrees however, the process of deriving the atmospheric density can be shortened by using equation IV defined as:

$$(IV) \quad e = 1 / \cos \theta$$

Before we go further, let us look at the range of values in the image representing the zenith angle.

- j) Use Metadata to view the characteristics of the image ZENANG. You will notice that the angles range in value between 6 and 10 degrees. Since the angles are less than 60 degrees, it is possible for us to use equation IV. Therefore, to compute the image showing the zenith angle for the various pixels, run TRANSFORM. Select the transformation option to produce the Reciprocal image. Specify ZENCOS (an image showing the cosine of the zenith angle, created earlier) as the input file and call the output DENSITY.

Atmospheric Turbidity (p)

Generally, atmospheric conditions for southern Chile in the summer month of February will include some clouds and therefore some turbidity. It will in this case be convenient to assume an atmosphere of maximum transparency and minimum turbidity.

- k) We will create an image to represent maximum insolation received at the surface. In this case, the image will have a value of 0.7. This is an estimated figure reflecting what proportion of the solar constant reaches the earth even under conditions of maximum transparency. Run INITIAL from the Data Entry menu. Specify the output image as TURBID and the image to copy parameters from as DENSITY. Select the output data type to be real and the initial value as 0.7.

Deriving Insolation Potential (IIP)

To complete the modeling of the Insolation Potential, we use equation I:

$$IIP = k p^e \cos v$$

In applying this equation, we need to raise the turbidity index to an exponent equal to the density index, then multiply the product by the solar constant and the estimated solar intensity.

$$IIP = 1.98 (\text{TURBID}^{\text{DENSITY}}) \text{INSOPOT}$$

- l) Run OVERLAY and specify TURBID as the first image, DENSITY as the second image, and call the output image DENTURB. Choose the exponential option (First to the power of Second). Next, run the module SCALAR to multiply DENTURB by 1.98 (the solar constant) and call the output image SOLAR. Finally, to obtain the Instantaneous Insolation Potential per pixel for the region for midday of February 11, run OVERLAY and multiply INSOPOT by SOLAR to create SOFINAL.

This image contains the potential insolation for the region. It is measured in units of langley's per minute per square centimeter. You may note that the shape of SOFINAL does not reflect the actual shape of the study area. An outline of the study area is contained in the image AREADEF. We will use it to mask out the region beyond the bounds of the study.

- m) Run OVERLAY and multiply AREADEF and SOFINAL. Call the output image FINAL. This contains the estimated insolation potential for the region.

2. *By how much is the estimated maximum insolation potential for the region greater or less than the solar constant (1.98 hy/min/cm^2)?*

Further Explorations

Finding Thermal "Islands"

The main objective of this exercise is to find an approximation of the thermal characteristics of the Puna Atacama region. It is believed that this will facilitate the identification of sources of heat concentrations identified as influencing the rainfall in the region. To this point, we have estimated the potential insolation for the region in langley/minute/centimeter units

of solar energy. We will examine the results relative to the elevation model. If we had additional data, we could compare the resulting surface against thermal satellite imagery using data from the same day.

- n) View FINAL and then CDEM in DISPLAY Launcher using the default palette. Do you find any relationship between the topography and the estimated insolation? It probably will not be possible to answer this question with FINAL in this form.

There are several steps that will improve the visual quality of FINAL. We might stretch the values over a range of 0-255. These values might then be reclassified into a set of value range categories which reflect heat or cold islands. To begin with, we will stretch the original image over 255 values.

The STRETCH module will rescale real number values to the 0-255 range. A simple linear stretch is the only option available for real number images. Before running such a stretch, it is necessary to examine the histogram of data values to anticipate any problems in doing a simple linear stretch.

- o) Run HISTO and specify the input image as FINAL. Enter a class width of 0.01, and use the defaults for the other options. The graphic histogram display is limited by the high frequency of zero values which we know represents the background. However, it is possible to see that relative to zero, the minimum data value is much higher, and the remaining range of values is skewed. This information is necessary to maximize the benefits of a stretched image.

To find the minimum value for use in the stretch, run the module EXTRACT and specify AREADEP as the feature definition image and FINAL as the image to be processed. Retain the default for finding the minimum value and select the output type to be tabular. This allows us to see the minimum value in the image, excluding the zero values masked out of the study area

3. *What is the minimum value?*

The maximum value is 1.4. We can now stretch the image.

- p) Now run the module STRETCH from the Display menu and specify FINAL as the input image and call the output image INSO__ST. Click on Leave out zero (as background value) from the input image. Next, click on Specify lower bound other than minimum and enter the minimum value as 1.1. Click on Specify upper bound other than maximum and enter the value as 1.4. Use DISPLAY Launcher and the Qualitative palette to examine INSO_ST.

Notice that the vast majority of pixels fall between 225 and 234. If we were to display all 256 color levels simultaneously using the Quantitative palette, these pixels would all be displayed in variants of green, and not be distinguishable easily by the eye. Autoscaling the image would suffer from the same problem because a skewed histogram results in a low contrast image (as we saw with FINAL). The question remains of how to best display this data to provide useful information.

One option is to reclassify groups of heat values into significant insolation ranges. To do this, we must determine what constitutes distinctly low and high insolation areas. By viewing the distribution of the heat values in the image, it is apparent that values cluster spatially. Exploring the value ranges in INSO_ST, a few things become apparent. Although values 1-212 vary greatly in temperature, they are spatially contiguous, and they suggest the presence of "cold islands." Values 212-234 cover very broad areas which one expects in low relief spaces. Values 234 and above tend to cluster also and suggest "heat islands".

Let us try reclassifying INSO_ST into a new file called ISLANDS. To verify the value ranges we determined above with our eyes, we may want to look at what percentage of values fall into the various ranges. To do so, we can use HISTO. However, HISTO will not be useful to us without first eliminating the background values in the image. We do this by first running QUERY.

- q) Run QUERY from the GIS Analysis/Database Query menu. Specify the input image as INSO_ST, the mask

image as AREADEF, and the output query image as AREA. A message may appear warning that the file is not for display. That is because we have eliminated the background pixels. Now run HISTO on AREA. Select the numeric output type. By looking at the cumulative percentage column on the far right, you will notice that the coldest bottom 5% of the values (excluding the 0 value pixels) are less than the value 173. The top 5% of the values fall between 229 and 234; these represent the heat island pixels. These breakoff points will be used in our reclassification of the image into separate categories.

- r) Run RECLASS on the image INSO_ST to produce the image ISLANDS. Reclassify the range of old values based on the above category ranges:

a value of 1 for all values ranging from 1 to just less than 173

a value of 2 for all values ranging from 173 to just less than 229

a value of 3 for all values ranging from 229 and up

- s) Use DISPLAY Launcher to view ISLANDS with the Qualitative palette. The pixels with value 3 are clustered together and represent heat islands. Next, we can examine the heat islands relative to the topography.

We can simultaneously compare heat islands with topography by draping the estimated insolation over the elevation model in an orthographic display.

- t) Run ORTHO from the Display menu. Specify CDEM as the surface image and ISLANDS as a drape image. Specify a view direction of 45, a viewing angle of 70, and keep 1 as the vertical exaggeration factor. Choose the Qualitative palette. Using other orthographic perspectives allows further investigation. With a low inclination view due North, the heat islands disappear, hidden behind the south facing slopes of the mountains.

4. *At what locations in the three-dimensional image are the high temperature concentrations found?*

Comparing the Modeled Results to Satellite Data

The question remains of how to verify the accuracy of this data. One means is to compare the information to satellite data sensed in the thermal range. To do so requires thermal imagery of the same date and time.

- u) Use DISPLAY Launcher and the default palette to view NOAAFEB.

As indicated earlier, this image represents the thermal characteristics of the region as recorded in the thermal band of the NOAA satellite. The data was collected on February 11, the same day as we have assumed for the exercise, and also has the same one kilometer spatial resolution as all other images acquired for this exercise. The data values are inverted so that higher values represent low temperatures and vice versa. Both FINAL and NOAAFEB contain temperature estimates representing a continuous relief surface and therefore can be compared.

- v) To compute the correlation index between the estimated insolation potential and the thermal characteristics recorded in the NOAA image, run the module REGRESS from the GIS Analysis/Statistics menu. Specify NOAAFEB as the independent variable and FINAL as the dependent variable, using NUBE4 to mask out the clouds in the NOAA image. Note the r , r^2 and the t-statistics.

You will probably find a very low correlation between the estimated insolation and the thermal image from the NOAA satellite for the total area. The generally low overall correlation is partly due to the fact that we assumed a condition of minimum atmospheric interference in the estimation of the surface insolation. Variables like clouds (recorded in the thermal image of the NOAA satellite) can distort the relationship among the temperature variables in both images.

More importantly, the insolation potential in FINAL is a modeled estimate of the amount of energy *reaching the earth*. The NOAA image on the other hand is a measure of solar energy transmitted back *from the earth* surface and recorded by a

thermal sensor. This is different from the estimated insolation in that thermal imagery senses emissivity. Emissivity is in part determined by the interaction of solar insolation with earth materials. To rectify this effect, it would be necessary to incorporate the emissivity characteristics of different land use types. Even so, some immediate comparisons are possible. Further work with these images revealed that while the overall agreement of the two images was poor, certain sub-regions of the images correlate well.

Conclusion

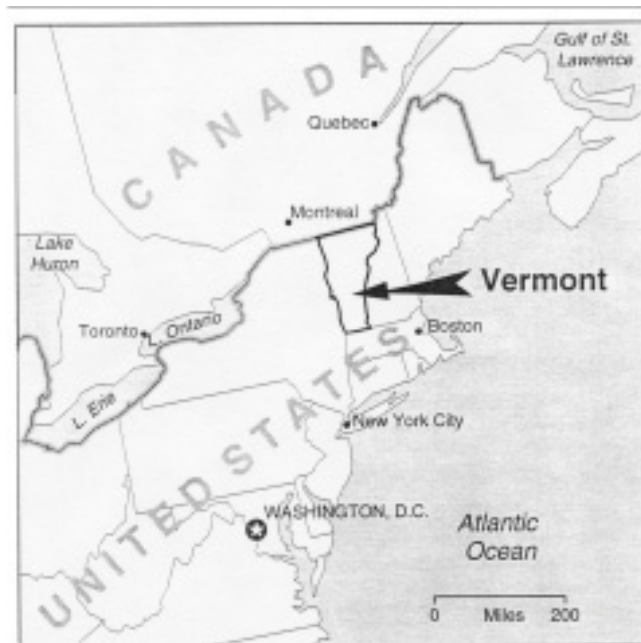
We have estimated the potential insolation intensity for the Puna Atacama region of southern Chile through the mathematical modeling of the atmospheric conditions and surface temperature characteristics of mountains in GIS. The result indicates that by using the equations based on principles of spherical trigonometry, it is possible to derive an approximation of the thermal conditions of a relief surface. Clearly, it is not easy to model highly variable conditions such as weather in environments stretching across several miles of rugged terrain. We must capture at a broad scale all of the parameters used at one moment in time. As in other fields of GIS, any kind of modeling based on a DEM requires a thorough understanding of the scale of research, the processes one can analyze at that scale, and the implications of the DEM scale on accuracy.

Procedures like this one might also be used to model terrain effects on other climatological phenomena at other scales. Microclimate modeling on broken terrain is a yet unexplored avenue for GIS application. With the obvious central role of solar input into physical and biotic phenomena on the earth's surface, models such as these will be sought by environmental, soil and hydrological scientists, and planners to estimate potential impacts of solar energy on the phenomena they study. In energy applications, insolation potential might also be modeled for the sighting of solar collectors or the organization of passive solar energy construction planning. Research in this field is in its infant stage and more work will be required.

References

- Barry, R.G., (1981) *Mountain Weather and Climate*. London: Methuen.
- Gimpel, S., (1987) *Insolación potencial en el espacio Geográfico*. Dcto. docente, Depto. de Geografía Universidad de Chile.
- Messerli, B. et. al., (1990) *Climate Change in the extreme arid Andes of Northern Chile*. Dcto. Universidad de Berna Suiza.
- Misetich, V., (1991) *El despoblamiento de los centros poblados precordilleranos de la IIª Región. Crisis cultural y colapso ecológico como factores detonantes del proceso*. Congreso de Geoecología de Montaña, Octubre 1991, Stgo, Chile.
- Rivera, A. y Romero H., (1992) *Aplicación de un Sistema de Información Geográfica al estudio Topoclimático del Norte de Chile*. XIV Congreso Nacional de Geografía, Talca, 1992.

Exercise 5: Mitigating Topographic Effects in Satellite Imagery: Case Studies of Vermont and Nepal



Introduction

Satellite image correction and interpretation is a complex procedure, even under the best of conditions. This is especially true for images from mountainous environments. The interaction of the angle and azimuth of the sun's rays with slopes and aspects produces a *topographic effect* resulting in variable illumination. This impedes our ability to make use of images in mountainous areas.

Topographic effect is defined simply as the difference in radiance values from inclined surfaces compared to horizontal ones. This orientation (sloping/flat) is measured relative to the angles of the sun and remote sensor (Holben and Justice, 1980). Images are often taken in the early morning hours or late afternoon, when the effect of sun angle on slope illumination can be extreme. In places such as the Green Mountains of the United States or the Himalayas of Nepal, reflectances of slopes facing away from the sun are considerably lower than the overall reflectance or mean of an image area. In highly mountainous terrain, some areas may be shadowed to the point that meaningful information is lost altogether.

Commonly, these effects manifest themselves during the classification procedure. Shadowing effects exaggerate the difference in reflectance information coming from similar earth materials. The same land cover type on opposite facing slopes may, under these conditions, be divided into separate classes. In the classification process, the highly variable relationship between slope, land cover, and sun angle can lead to a highly exaggerated number of reflectance groups that make final interpretation of data layers more costly, difficult, and time consuming. This is only the simplest of possible scenarios. While very often land covers are *not* the same on opposite sides of a mountain since, to a certain extent, slope and aspect contribute to these patterns (see Chapter 6), shadowing nonetheless makes it difficult to derive biomass indexes or other comparisons among land cover classes. For these reasons, the effect of relief on image interpretation can be dramatic and has been the subject of a great deal of research in GIS, Remote Sensing, and Photogrammetry.

These problems are caused by the persistent situation of sensors operating when sun elevation levels are low. In their research, Proy et al. (1989), concluded that, given the computational level required for correction, researchers are better off accessing images taken with a higher sun elevation, possibly greater than 45, than attempting to work with low sun elevations. Because of the limits on data availability, this is not a realistic option however. Researchers must find ways to make use of whatever data may be available.

As a result, it is necessary to use techniques that might improve the amount of information that we can derive from such images. Several such techniques have evolved over the years. In order of increasing sophistication, these techniques include: the *ratioing* of different bands of information, the *partitioning* of an image into separate areas for classification, the *modeling* of illumination effects from direct radiance using a DEM, and the *calculation of backwards radiance* and indirect diffusion effects. This exercise seeks to introduce these various methods and to put the first three of them to use on sample data¹.

Band Ratioing

In band ratioing, one band image is divided by another. The infrared values, for example, are divided by the red. This may be a simple division:

$$\text{Band A} / \text{Band B}$$

or often, a more robust normalized ratio:

$$(\text{Band A} - \text{Band B}) / (\text{Band A} + \text{Band B})$$

The latter is employed to limit the value range and avoid division by zero. The resulting output image is then used for further clustering or other analysis. Band ratioing is based on the principle that a certain component of reflectance in all bands is a result of angular effects. These effects are experienced as a uniform multiplier to reflectance values that are oth-

1. Data and a demonstration of the exercise was contributed by Thomas Millette, Department of Geography, Mt. Holyoke College, South Hadley, MA, USA.

erwise determined by real differences in earth materials. By dividing one band by another, the uniform angular component is divided away, leaving the variation which represents differences in earth materials (Holben and Justice, 1981). Variations resulting from atmospheric effects, scattering, and other anomalies may remain, however.

Historically, the bands chosen for such an analysis have been the Infrared and Red ones². This is in part because for many earth materials such as vegetation, the variation between these bands is maximal. In addition, many of the often-significant additive effects to reflectance values (like atmospheric scattering) are experienced similarly in these two highly proximate bands. The ratio between them therefore does not produce artificial variance in the output image (Holben and Justice, 1981). Other band ratios are possible. Ratios of the values of MSS Bands 6 and 5 as well as 5 and 4, for example, produce different ranges and variations (Justice et al., 1981). However, for the goal of controlling topographic effects (rather than highlighting various specific earth materials), Infrared and Red continue to be the most commonly employed. Although it is one of the simplest and earliest developed techniques for the mitigation of terrain effects, band ratioing continues to be a widely used and effective method.

Band ratioing does not fully eliminate or control the topographic effect however. Many deeply shaded areas and high relief areas in satellite images, having very low values across several bands, will be unimproved through band ratioing. The effect of scattered and diffuse light is similarly unimproved.

Image Partitioning

Image partitioning works from the simple assumption that since different areas within an image are affected differently by illumination effects resulting from slope and aspect, these distinct areas therefore should be classified separately. To conduct this correction, the image is separated into several smaller images and divided based upon elevation, slope, and aspect, or simply upon the evidence of heavy shadowing. These subscenes are each clustered separately and then put back together as a final classified image.

Image partitioning is another simple and somewhat effective technique. However, the criteria used for the division of the image are crucial. The resulting clustered image may suffer from arbitrary separation along slope and aspect lines which do not reflect significant thresholds for the topographic effect. These thresholds are determined not simply by the slope and aspect of the surface, but by the inclination of the surface relative to the angle of the sun's radiance. Modeling this phenomenon requires more sophisticated procedures.

Further, the introduction of a DEM to the procedure of image correction and classification increases the possibility of error and uncertainty. The registration of the DEM to the satellite data must be precise, otherwise such a correction may do more harm than good.

Modeling Direct Illumination Effects Using a DEM

According to a simple Lambertian assumption, the portion of radiance reflected off of the surface of a plane is a function of the slope and aspect of the plane relative to the slope and azimuth angle of the sun's radiance at a given moment. This relationship is described in spherical trigonometry (Smith et al., 1980) by³:

$$\cos \nu = \cos \theta \cos \alpha + \sin \theta \sin \alpha \cos$$

where

θ = zenith angle of the sun

α = slope of the plane

2. Infrared and Red bands are bands 7 and 5 from MSS, bands 4 and 3 from TM, and bands 3 and 2 from SPOT scanners respectively.

3. For more discussion of this relationship and its modeling in GIS, see Exercise 4 in this volume.

d = difference between the azimuth of the plane and the azimuth of the sun

$\cos\theta$ is the proportion of real reflectance coming off a slope face. By using a DEM to derive slope and aspect, the equation can be solved using map algebra in GIS to create an output image of $\cos\theta$. In theory, "true" radiance images can be created by dividing the band images by $\cos\theta$ (Smith et al., 1980). These "corrected" bands may then be used to create composites and used for classification, or ranges of $\cos\theta$ may be used to partition the image for compositing, clustering, and classifying as described above.

This technique has a number of limits as well. In practice, the method was found to overcorrect the data, causing wider variation for many land use classes than the uncorrected data (Justice et al., 1981). Additionally, it has been shown that the Lambertian assumption underlying the correction was valid for only a limited range of sun and slope angles (Smith et al., 1980). Finally, the technique does nothing to account for data variation caused by diffuse and indirect illumination resulting from topography.

In any case, a high quality and well-registered DEM is necessary for this technique. In many parts of the world, the quality of elevation models is limited by the availability of detailed elevation data used for the interpolation of a surface (see Exercise 1). The geo-registration of the DEM is crucial as well, and disagreement between image and DEM registration will inevitably lead to poor results. Tests of this technique have also shown, among other things, that it is important to have a DEM of the same resolution as the image being corrected and that the degree of the required correction differs between wavelengths⁴.

Scattering Effects and Non-Lambertian Assumptions

Correction techniques which incorporate non-Lambertian assumptions to account for scattering effects and indirect illumination are the most sophisticated and experimental. Colby's (1991) application of this technique covered only a very small subset of a satellite scene and still requires more testing and confirmation. The principle behind the approach is that much of the illumination enhancement or dampening in a region is the result of *indirect* radiation, light which is scattered off of other earth surfaces and material in the atmosphere before bouncing into the observed land surface and returning into space and the lens of the observing satellite. The correction previously explained uses a DEM to model the possible effects of slope and aspect only on the *direct* radiance from the sun to the surface and back.

This indirect scattering effect is often referred to as the *bi-directional reflectance distribution function*. It is incorporated into a model which resembles other DEM-based corrections except that it uses a non-Lambertian assumption. This means that it derives and employs a locally-derived constant which reflects "surface roughness."⁵ Like other DEM based corrections, a wealth of error may be introduced through the use of anything less than a high quality elevation model with excellent geocorrection. Owing to the complexity and experimental character of this technique, we introduce it here without an application example.

The Exercise

Three methods for mitigating topographic effects on satellite images are presented here. The first case study comes from the Dhading District of Nepal and shows the use of spectral band ratioing. The second case study is from Bennington, Vermont in the United States and shows the analysis of a partitioned image. The last example uses the Bennington data to examine the use of a DEM in modeling illumination effects for mitigating shadow effects using spherical geometry.

The band ratioing and image partitioning steps outlined here reflect standard protocols at many geoprocessing labs but

4. Civco (1989) employed a more sophisticated example of this technique, involving a two-stage approach. A DEM was first used to model topographic effect as described above. This was followed by a second correction which utilized an empirically-derived correction coefficient based upon the differences between image reflectance values for various known earth materials.

5. The Minnaert constant (Smith, 1980).

do not represent a universal prescription. We choose these techniques for their relative simplicity, but also because given the lack of ancillary data in many parts of the world, especially good elevation models, it seems appropriate to demonstrate methods most commonly used. There is no universally recognized solution to the problem of uneven reflection, and currently, techniques vary according to the variables they incorporate. These techniques continue to be, for the most part, experimental.

Band Ratioing

We first use a simple band ratio of Landsat TM bands to illustrate the effects of relief shadowing on satellite imagery. This is a common method for minimizing the differences in brightness due to the effects of topography. The ratioing of image bands is a process whereby a pair of images (each representing a different spectral band at a given location) are divided to create an output image showing the ratio of values between them (e.g., Band A / Band B). In the output image, the ratio values for the same earth object will tend to approximate each other, thus removing the variation in reflectance resulting from the topographic effect. Consequently, the process is useful to reduce the more severe effects of shadowing on satellite images.

The data for this part of the exercise consists of NEPC2, NEPC3, and NEPC4. These are raw TM images from bands 2, 3 and 4 covering parts of the Himalayas in Nepal.

- a) Use DISPLAY Launcher to view the images NEPC2, NEPC3, and NEPC4 with the GreyScale palette. These are the TM Green, Red, and Infrared bands usually used in creating a false-color composite and subsequently clustered for analysis.
- b) Run HISTO with each band to examine the distribution of the reflectance values. Choosing a graphic output, the shapes of the value distributions are clear. The histogram distributions show poorly contrasted information. In all three images, the information is concentrated in the middle values while little information occurs at the extreme ends. By running HISTO again and choosing numeric output, we see that the small top and bottom 1% of the data values span long ranges at the ends of the histogram.

1. *According to the histogram, what are the current minimum data values for each of the three images? What are the approximate maximum values of the bottom 99% of the data in the three images?*

To enhance the contrast in the image for display purposes, we must run a saturated contrast stretch. This will serve to push the minimum and maximum limits of the histogram across a full range of data values.

- c) To do this, run STRETCH three times, choosing the Linear with Saturation option to rescale NEPC2, NEPC3, and NEPC4 to 256 levels. Accept the default for a 1% saturation. Call the output images NEPC2S, NEPC3S, and NEPC4S. View the three images in DISPLAY Launcher with the GreyScale palette.

2. *Explain why the images display with a better contrast than before stretching.*

These images are clearer for viewing and tell us that a 1% saturation increases the amount of contrasted information available. We will now use the three images in a composite.

- d) Run the module COMPOSITE from the Display menu and specify NEPC2, NEPC3, and NEPC4 as the Blue, Green, and Red bands respectively, and call the output image COMP1. For the contrast stretch type, choose Linear with saturation points, and accept all other defaults. Click OK.

Next, we will produce a ratioed composite image for comparison with image COMP1. While ratioing minimizes brightness effects, it also tends to be "intensity blind" in that different earth materials may have similar slopes in their spectral reflectance curves, and so may appear identical in the ratioed image. Therefore, we will use only one ratio image in the composite along with two other "raw" bands. We will take the ratio of Band 4 and 3 and use the result image as a substitute for Band 4 in a 234 composite. Bands 3 and 4 were chosen because of their spectral proximity (as discussed above)

and because Band 4 contributes the most information of the three bands⁶. We will use a normalized ratio $(A-B/A+B)$ since it is more robust than a simple ratio and disallows division by zero.

- e) Run OVERLAY and specify NEPC4 as the first image, NEPC3 as the second, and call the output NRATIO. Select the overlay option for normalized ratio (First-Second/First+Second). Use Metadata to examine the documentation file of NRATIO.

3. *What are the minimum and maximum data values of NRATIO?*

We will need to transform the range of values in NRATIO to a byte range (0 to 255). We might use map algebra to add to the values to reach the positive range and then multiply to scale the values between 0 and 255. A simple linear stretch will work just as well, however.

- f) Run the module STRETCH and specify NRATIO as the input image and NRAT as the output. Viewed in DISPLAY Launcher, NRAT should look close to NRATIO.
- g) Next, make a second composite, this time using NRAT as one of the components. Run COMPOSITE to combine NEPC2S, NEPC3S, and NRAT as the Blue, Green, and Red bands. Here, we use the stretched input images and so saturation of the ends of the histogram is unnecessary. Choose only a simple linear stretch for the composite. Call the output image COMP2.
- h) Look at COMP2. You will notice that vegetation in the shadows is more visible than before. Some vegetation appears distinguishable enough to classify into specific categories.
- i) Compare COMP1 and COMP2 using DISPLAY Launcher.

4. *How does the ratioing of the images help in visual interpretation?*

In cases of very steep slopes where low sun elevation is more likely to cause extremely deep shadows, a band ratio is one of the better options available. Its application can be problematic, however. While it normalizes the shadowed slopes and improves discrimination of land cover in these areas, the method tends to exaggerate biomass everywhere else. In the process, we can lose discrimination between different land cover types as the range of reflectances in the more illuminated areas decreases.

Given that whatever vegetation grows in shadowed areas is probably controlled to some extent by these shadow effects, other correction options are available. In particular, image partitioning may be effective where cover types mimic variations in slope and aspect.

Image Partitioning Using Slope and Aspect

As discussed above, another solution to the problems created by the topographic effect is the partitioning of an image into separate areas and the subsequent clustering and classification of the separate areas. The criteria for this subdivision should be based upon the likely combinations of slope and aspect which may result in illumination effects.

A simple version of this technique might not require the use of a DEM. Using the composite or a single band, the deeply shadowed areas of an image may be separated out, using onscreen digitizing. These digitized polygons, representing areas of deep shadow, may then be made into a mask and used to isolate images of shadowed and non-shadowed areas. These may then be independently clustered and classified into a final compound image. The division of shadowed and non-shadowed areas is somewhat arbitrary however, and given to inaccuracies.

Blamont and Mering (1985) present a technique that segments data into three slope categories, clusters them, and treats

6. One can confirm this by running an unstandardized principle components analysis on all the bands using the PCA module.

classification of the data separately for each category. Like band ratioing, the technique is intuitive and simple, and can serve as the basis for more complex rectification techniques. This exercise compares the results of an unsupervised classification on an unrectified standard color composite and another false color composite divided into slope and aspect categories before clustering. Finally, the clustered images are compared to the result of a supervised classification. The data consists of the following:

- 1) BENCOMP, a false color composite of TM bands 2, 3, and 4 covering parts of Bennington, Vermont, USA;
- 2) BENCLUS1, an unsupervised classification from the TM bands
- 3) MAP3, a supervised classification of the TM composite bands; and
- 4) BENDEM, a digital elevation model of the same area.

The first part of the exercise is a simple clustering of the composite values.

- j) Let us examine the false color composite image BENCOMP. Use DISPLAY Launcher and the GreyScale palette to view BENCOMP. The right side of the image is mountainous while the left side is a broad valley bordered by some hills. Notice the shadowing effects in the mountains.
- k) Look at BENCLUS1 with DISPLAY Launcher. This is an unsupervised classification image of the study area. In this case, 16 significant clusters were identified.
- l) A classification of these clusters is supplied for you in an integer values file called BENVAL1. This collapses the clusters into information categories based on some field reconnaissance. Use Edit to look at BENVAL1. Then run ASSIGN. Specify BENCLUS1 as the feature definition image, BENVAL1 as the values file, and MAP1 as the output image. MAP1 is the final non-topographically corrected, clustered land cover map.
- m) Open Metadata for the image MAP1. Double-click on the Categories option and add the legend categories with the following information:

1	hardwood
2	softwood
3	crop fields
4	hay/pasture
5	developed
6	water/wetland
7	cloud

Save the changes and redisplay MAP1.

Provided in the data set is an image called MAP3. This image is the result of a supervised classification and extensive ground reconnaissance. While it is not a perfectly corrected image, for general purposes we may use it as a control to compare our roughly clustered image.

- n) Display MAP3 with the user-defined palette called MAP3PAL.
5. *Compare MAP1 and MAP3 in terms of landuse category coverages. What category is distinguished in MAP3 that*

does not appear in the clustered image? Why might this category have been lost in the original clustering analysis?

Clearly, the topographic effect is hampering the clustered classification of the satellite data. The next part of the exercise uses a DEM to help control this effect. In this case, we will partition the composite image into slope and aspect categories that are then clustered separately.

- o) Look at BENDEM with DISPLAY Launcher. Next, run SURFACE and select to create slopes (in percent) from BENDEM and call the output BENSLOPE. Run SURFACE again, this time to create an aspect map from BENDEM, and call the output BENASP. Run DISPLAY Launcher to examine both images. Notice that flat areas are given an aspect value of -1.

By knowing the general direction and sun angle at the date and time the satellite passed over the study area⁷, we can develop a set of slope and aspect categories. Based on the general angle of illumination, it was decided that the areas of differential illumination were low slope areas, higher slope areas facing north and west, and higher slope areas facing south and east. The following classes represent those divisions:

- | | | | |
|----|-------------------------------------|---|---------|
| 1) | slopes < 10% | = | LSLOPE |
| 2) | north and west facing slopes > 10 % | = | NWSLOPE |
| 3) | south and east facing slopes > 10 % | = | SESLOPE |

We need to create three separate Boolean images that represent each of these categories.

- p) We can use a macro file to create the three images and save time and effort in performing the individual operations. Run Edit and open the macro file titled SLOPES. The operations required to create the three images are contained in this macro file. Go to the File menu and select Run Macro. Specify the name of the macro as SLOPES. After the macro has finished, use DISPLAY Launcher to look at each of the three Boolean images. Notice that each pixel belongs to only one of the three groups.

Now we are ready to create three images whose reflectance values are based on the slope and aspect layers of the relief in the study area. We do this by masking out those areas in the composite image, BENCOMP, by applying the three images we just produced.

- q) Run OVERLAY to perform the following operations:

LSLOPE x BENCOMP = COMPLOW

NWSLOPE x BENCOMP = COMPNW

SESLOPE x BENCOMP = COMPSE

- r) The images CLUSLOW, CLUSNW, and CLUSSE have been provided for you as a result of running CLUSTER.

Each of these images has categories corresponding with the various earth materials in the area. However, since CLUSTER assigns values to the clusters based upon the size of the area covered in each image, there is no way to know that a value of 1 in CLUSLOW is the same material as a value of 1 in CLUSNW or CLUSSE. We must give each cluster type a unique identifier. We do this by simply adding a constant to each of the images.

- s) Use Metadata to view the documentation files for CLUSLOW, CLUSNW, and CLUSSE. We find they have 12, 12, and 15 clusters respectively.

7. This is usually found in the header of the original satellite data file or in accompanying documentation.

Image Name	Number of Clusters	Desired Cluster Value Range	Value to Add
CLUSLOW	12	1 through 12	0
CLUSNW	12	13 through 24	12
CLUSSE	15	25 through 39	24

- t) Run SCALAR twice and add a constant of 12 to CLUSNW to create CLUSNW2, and a constant of 24 to CLUSSE to create CLUSSE2.

In order to overlay these images on top of one another, we must finally reclassify the clusters representing their background value back to 0, to permit a covering operation to combine them.

- u) This is done simply by running RECLASS twice. First, run RECLASS on CLUSNW2, calling the output file NORTHW, and give a new value of 0 to old values ranging from 13 to just less than 14. Then run RECLASS again on CLUSSE2, creating a new image called SOUTHE, giving a new value of 0 to all old values ranging from 25 to just less than 26.
- v) We must finally overlay the partitioned pieces of the image back together. Run OVERLAY twice and choose the cover option both times.

```
COVER      SOUTHE      over      NORTHW      = TEMP
COVER      TEMP        over      CLUSSE        = BENCLUS2
```

- w) Look at BENCLUS2 with the Qualitative palette. Notice the large number of clusters created through this process.

The final step involves classifying clusters into meaningful categories. Following extensive ground reconnaissance, this was done, producing the image MAP2.

- x) Look at MAP2 in DISPLAY Launcher applying the user-defined palette called MAP2PAL. A classification of the land cover categories based on field reconnaissance is provided in a values file called MAP2VAL. By running ASSIGN with this values file on BENCLUS2, MAP2 was generated.
- y) We now have two classified images, one based on unpartitioned input images (MAP1), the other divided along general slope and aspect categories (MAP2). View each of these in turn with DISPLAY Launcher to compare them. Notice the improvement in discriminating classes in the shadowed areas in the eastern part of the image.

To compare MAP2 with our control MAP3, we may use a crosstabulation.

- z) Run CROSSTAB and specify MAP2 as the first image and MAP3 as the second image. Create both crosstabulation and an image. Call the output image CROSS1. Ask for a Kappa Index of agreement. Then look at CROSS1 in DISPLAY Launcher in order to compare the two maps.

6. *How does MAP1 compare to MAP2. How well do they agree? What accounts for this difference?*

You will notice that there is not perfect agreement here and that areas of disagreement dominate the mountainous and shadowed areas. The massive effort required in groundtruthing and making sense of the much expanded list of clusters created through partitioning, in no way ensures a problem-free classification.

Depending on the nature of mountains and sun illumination at the moment the sensor is used, partitioning and clustering of images may create more problems than it solves. Such partitioning works best where land cover conditions are stratified environmentally. Otherwise, the potential to create hundreds of meaningless clusters arises.

A more fundamental problem lies in the arbitrary and inexact nature of the thresholds for slope and aspect based division of the image. A more precise algorithm, which calculates the angle difference between the slope/aspect of a plane and the illumination of the sun, may be used to more precisely model the illumination effects using a DEM.

Modeling Illumination Effects with a DEM

The trigonometric formula provided in the introduction (used and more fully explained in Exercise 4) describes the angular relationship between the slopes and angles of a surface plane and the sun. By using input images of slope and aspect derived from a DEM as a substitution for α and d in the equation, along with values for the angle and aspect of the sun, we can solve for the coefficient $\cos\theta$ using map algebra.

In this case, we will use DEM data from the high relief Njolomole region of central Malawi.

- aa) Use DISPLAY Launcher to look at NJOLODEM. The dramatic relief of the escarpment in the west overlooks the Great Rift Valley to the east. Though not provided here, satellite images from all bands and the composite images created from them, show deep shadowing effects along this rift line.

To attempt to compensate for this topographic effect and draw out some of the lost values from the deeply shadowed region, we might want to subdivide the image between areas of high and low slope. This method may provide deceptive results, however. Depending upon the angle and slope of the sun at the time a satellite image is taken in this region, the area experiencing topographic effect may change. To determine the areas where separation or compensation are required, we must include not only slope and aspect data from the surface, but information about the relative angle of that surface to the direct illumination of the sun; we must calculate the angular deviation of the sun from every pixel. To do this, we will follow the map algebra steps laid out in the spherical trigonometric formula given above.

These map algebra steps involve the use of the IDRISI modules TRANSFORM, SCALAR, and OVERLAY. A macro file has been provided here which covers the commands in IDRISI which use the slope and aspect images to solve for this coefficient⁸. The commands follow this outline:

- 1) Slope and aspect values are calculated from NJOLODEM.
- 2) These are transformed from degrees into radians.
- 3) The sine and cosine values of these angles are taken.
- 4) The difference in the sun's aspect (37.73 degrees) and that of the image pixels is calculated and the absolute value of the difference is taken.
- 5) The cosine of the difference is calculated.
- 6) The cosine of the sun's angle (22.5 degrees) and that of the slope's angle are multiplied.
- 7) The sine of the sun's angle and that of the slope's angle are multiplied.
- 8) The product is multiplied by the image created in Step 5.

8. These steps mirror those used in Exercise 4 and are discussed in greater detail there. In this case, the value of the sun's angle and azimuth were calculated for a satellite image taken at 9:46 am local time. The zenith angle of 68.5 degrees and the azimuth of 37.73 degrees were found using the VOYAGER[®] desktop planetarium software package. These variables were entered into the formula as constants assuming a uniform sun angle and azimuth for all pixels in the satellite image.

- 9) The products of Steps 7 and 8 are added.
 - 10) The absolute value of this image is taken to create an image called COSV.
- ab) Run Edit and open the file named LAMBERT.IML. The steps outlined above have been operationalized in this file. Note that the constants used here are the values of the sine and cosine of the sun's zenith and azimuth (in radians) at the time of the imaging.
 - ac) Next, go to the File menu and select Run Macro. Specify LAMBERT as the macro file. After the macro has finished, use DISPLAY Launcher to look at COSV.

These are the absolute values of $\cos\theta$ in the study area. They range from 0 to 1 and represent the proportion of radiance reflected from the surface based upon the Lambertian assumption. Low values (near 0) reflect a high topographic effect. High values (near 1) receive near-direct radiance and require little correction. To examine the correction relative to the topography, we can drape one over the other. We must first stretch the correction image into a 16 value integer format.

- ad) Run STRETCH on COSV. Choose a simple linear stretch. and call the output COSV2. Enter 16 as the number of levels. Next, run ORTHO with NJOLODEM as the surface image and COSV2 as the drape image. Select the Quantitative palette for display and take the other defaults. After the image is displayed, you may need to turn on Autoscale (from Composer) to see it clearly.

You will notice that some of the areas of high slope which fall along the edge of the rift valley have low values while others do not. Not all areas of high slope experience a high illumination effect. Where the angle faces the sun, high slope pixels may receive near-normal radiance.

Ideally, the dn values in the band images for the area should be divided by these values to produce a corrected set of images for compositing and classification. However, this technique has been shown to cause dramatic overcorrection. In the resulting image, the values of COSV control too greatly the output dn values (Jones et al., 1988; Kawata et al., 1988). One possible solution is to recalculate the $\cos\theta$ values using a new slope angle where the input surface slope value (α) is multiplied by the sine of the solar zenith angle (following Ciccone, 1977). This may still result in overcorrection. As a *direct* correction coefficient, $\cos\theta$ may be inadequate.

These values $d\theta$, however, give an accurate measure of the deviation of the sun's angle from normal for every pixel in the image. They therefore provide an excellent basis from which to subdivide the image based upon objective criteria. To do this, we would reclassify COSV into three Boolean subdivisions based upon their $\cos\theta$ values and use these masks to separate and recluster the false color composite created from the raw dn values as we did previously. By modeling the illumination angle of the sun at the time a satellite image is taken, objective criteria can therefore be established for subdividing and clustering images taken in areas of high relief.

Conclusion

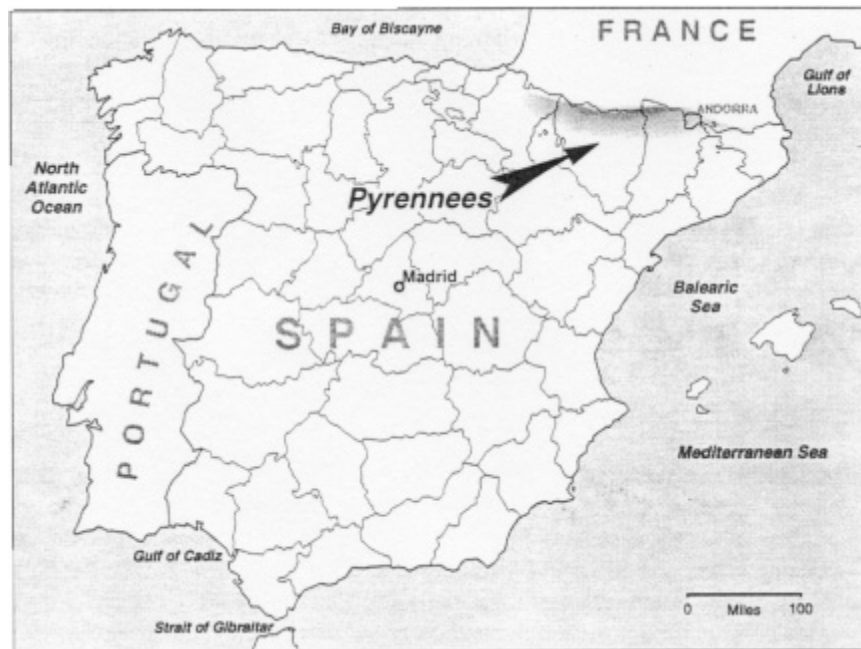
Despite the long history of development of techniques for the mitigation of terrain effects on satellite images, much of the work is still experimental. Provided here are some of the steps used by many labs as the standard protocol in the processing of images from areas of high relief. Other methods are available. More complex algorithms accounting for non-Lambertian assumptions may be applied. It is also possible to incorporate atmospheric effects into image corrections (Kawata et al., 1988). These atmospheric effects are often a function of relief and require the use of elevation data. The same caveats concerning the use of DEMs apply to these more sophisticated solutions.

More simply, the ratioing of different band combinations might provide good results in some areas. The traditional 4/3 ratio is a recommended but by no means universal proscription. Another possible method might involve the use of Principle Components Analysis (PCA) on several bands of input information from a mountainous area. In theory, the areas deeply affected by shadowing might be removed as an individual component. It is our hope that by presenting a demonstration of these simple techniques, we give GIS users some beginning tools to experiment with in their own contexts.

References

- Blamont, D. and C. Mering (1985) "Use of Remote Sensing for Vegetation and Landuse Mapping in Mountainous Area: The Case of Central Nepal," (mimeo) Kathmandu, Nepal.
- Cicone, R.C., W.A. Malila, and E.P. Crist (1977) "Investigation of techniques for inventorying forested regions" *Forestry information system requirements and joint use of remotely sensed and ancillary data. Final Report*. Volume II. NASA- CR- ERIM- 122700-35-F2.
- Civco, D. L. (1989) "Topographic Normalization of Landsat Thematic Mapper Digital Imagery" in *Photogrammetric Engineering & Remote Sensing*, 55(9): 1303-1309.
- Colby, J.D. (1991) "Topographic Normalization in Rugged Terrain" in *Photogrammetric Engineering & Remote Sensing*, 57(5): 531-537.
- Holben, B.N. and C.O. Justice (1980) "The Topographic Effect on Spectral Response from Nadir Points" *Photogrammetric Engineering & Remote Sensing*, 46(9):1191-1200.
- Jones, A.R., J.J. Settle, and B.K. Wyatt (1988) "Use of Digital Terrain Data in the Interpretation of SPOT-1 HRV Multi-spectral imagery" *International Journal of Remote Sensing*, 9(4):669-682.
- Kawata, Y. S. Ueno, and T. Kusaka (1988) "Radiometric Correction for Atmospheric and Topographic Effects on Landsat MSS Images" *International Journal of Remote Sensing*, 9(4):729-748.
- Leprieur, C. E., J.M. Durand, J.L. Peyron (1988) "Influence of Topography on Forest Reflectance Using Landsat Thematic Mapper and Digital Terrain Data" in *Photogrammetric Engineering & Remote Sensing*, 54(4): 491-496.
- Proy, C., D. Tanre, and P.Y. Deschamps (1989) "Evaluation of Topographic Effects in Remotely Sensed Data" in *Remote Sensing the Environment*, 30: 21-32.
- Smith, J.A., T.L. Lin, and K.J. Ranson. (1980) "The Lambertian Assumption and Landsat Data" in *Photogrammetric Engineering & Remote Sensing*, 46:1183-89.
- Smith, L. (1980) *Ecology and Field Biology*. 3rd Edition. Harper and Row: New York.

Exercise 6: Spatial Relationships between Ecosystems and Topography: Modeling Landscape Ecology



Introduction

The introductory paper to this volume summarizes the nature of the landscape approach. Drawing on the principles of landscape ecology, landscapes can be understood and modeled based on their constituent parts and interrelations. In landscape ecology, as in any form of simulation modeling, we may apply multivariate statistical analyses to evaluate the causal factors of a phenomenon or to predict a phenomenon's distribution. In research, this form of analysis may require the use of a number of packages, both statistical and geographical.

This exercise presents a study from the Izas catchment in the Pyrenees mountains of Spain. It employs multiple packages, including GIS, to model the contribution of various *topographic conditions* to resulting *landscape types*. Topographic conditions here include slope, elevation, drainage, and a number of other variables, while landscape types represent grouped combinations of vegetation, landform, and soil. Topography is here treated as the independent variable, determining the outcome landscape to a greater or lesser degree. The objective of the study is to determine the extent to which various limiting topographic factors explain the surface landscape in a location and throughout a region. Using this information, geographic images of predicted landscape distributions can be created using topographic input data in combination with sampled point data sets from the landscape being modeled.

The introduction to the exercise takes us through the preliminary steps used to build the data set. The second part of the exercise explains the use of Binary Discriminant Analysis for analysis of the topographic variables involved. The next section uses a batch file to create a set of Boolean images for use in the GIS component of the analysis. The final part of the exercise uses the Boolean images along with statistical tests against the original sample data to build maps of landscape groups.

The General Procedure

The general procedure for this form of modeling occurs in several steps; GIS represents only the final analytical operation in a more lengthy process. The method begins with the sampling of a wide distribution of points throughout the mountainous study area and the recording of various ecological variables at each location. This is followed by a statistical analysis of the interrelation of these variables. This results in a statistical measure of the degree to which the presence or absence of various topographic variables (slope, aspect, drainage, etc.) explain the presence of landscapes types (based on vegetation, soil, and geomorphology). Using separate images of these *topographic variables* and information about their hierarchical contribution to different *landscape types*, images of landscape group distribution are then created. For more detailed discussion of field methods, data development, difficulties, and results, please see the original researcher's publication¹.

This exercise will focus on the GIS application which finalizes the research. A description of the steps preceding the GIS methodologies is provided here. Briefly, the approach is as follows:

- 1) 237 points were sampled throughout a catchment in the Pyrenees mountains of Spain. At 194 of these points, landscape and topographic data were recorded. This included the recording of a variety of attributes which help describe horizontal heterogeneity -- vegetation, soils, geomorphology, regolith thickness, and origin of the substratum.
- 2) Separate landscape groups were defined by means of numerical taxonomy techniques on the field observations. These serve as the dependent landscape variables to be explained by various topographic characteristics.
- 3) Statistical methods were used to detect and order a hierarchy of relationships between the landscape groups and the topographic variables. Each independent topographic variable was given a set of values which reflects its contribution to the various landscape groups.

1. The exercise is based on the doctoral dissertation by Gabriel del Barrio Escribano, *Respuesta Topográfica del Paisaje en Alta Montaña: Pirineos Centrales*, presented at la Universidad Autónoma de Madrid, la Facultad de Ciencias, Sección Biológicas, June 1992. This exercise is a merging of his draft of the exercise, and his English translation of chapter 4, "Topographic control of the distribution of ecosystems," from the dissertation.

- 4) GIS images were created for each topographic variable.
- 5) The relationships between the topographic variables and the landscape groups were modeled using GIS to produce maps of potential distributions of landscape groups.
- 6) The maps were overlaid to examine spatial relationships between the landscape groups.

Field Sampling

The study area is the Izas catchment of the Central Pyrenees. The ecological attributes of 237 points were recorded and input into the software package TWINSpan to create vegetation classes (Hill et al., 1975; Hill, 1979). Soils were classified by their features according to FAO-UNESCO classification (1985). Geomorphic types were the result of taking into account landforms at each sample point.

Defining Ecological Groups

The next step was to classify vegetation classes, soils, and geomorphological results into ecological groups. A number of possible methods might be used to group the independent observations into significant groups. In this case, a binary matrix of the vegetation, soils, and geomorphic classes was produced for all the sampled points. This matrix was used to carry out the classification with a hierarchical agglomerative clustering method. This uses a hierarchy of attributes for the grouping process. Similarity was determined using the Jaccard index which is a co-occurrence measurement (Sneath & Sokol, 1973; Jackson et al., 1989). All procedures for this were carried out with the package SYN-TAX (Podani, 1988).

The result was to create ten significantly distinct landscape groups. Each of the 237 points was then assigned to these groups, based upon the combination of landform, soil, and vegetation at that site. Leaving aside some of the more technical characteristics, the ten groups included the following:

- Group 1: Well developed, thick soils; frost action; dense cover; varied species.
- Group 2: *Festuca eskia* and *Nardus stricta* dominated; dense cover; concave terracettes and hollows.
- Group 3: Dense grasslands; thick, well drained soils.
- Group 4: Like Group 2; more open cover; *Trifolium alpinum* is a regular component.
- Group 5: Grassland; slow draining; snow-fed moisture.
- Group 6: Dense grassland; water accumulation; dominated by *Nardus stricta* and *Carex nigra*; very thick soil.
- Group 7: Hydromorphic soils; continuous cover.
- Group 8: Open formations of *Festuca eskia*; turf-banked lobes and terracettes.
- Group 9: Thin lithosols or regosols; *Festuca eskia* and *Nardus stricta* codominant.
- Group 10: Like Group 9; *Festuca eskia* dominant and soil is ranker.

These were then analyzed against independent topographic variables using Binary Discriminant Analysis.

Discriminant Analysis of Topographical Relations: Non-Spatial Analysis

The next step was to examine the statistical relationship between the topographic variables and the landscape types based on the data from the field survey. This form of analysis was strictly statistical and involved no use of GIS techniques. For

this procedure, Binary Discriminant Analysis was employed to examine the relations between the groups and topographic variables². The goal was to determine the relative importance of each topographic variable to each landscape group, then use the results to map the distribution of each group. To explain group distributions according to several variables, it was necessary to adopt a multivariate approach. The binary approach was used because of the qualitative nature of the dependent variables.

Simply, BDA technique evaluates the presence or absence of the topographic variables against the presence of the particular landscape type. The method combines quantitative data with qualitative data. In this case, quantitative measures of topographic variables are turned into simple binary variables. A topographic variable is considered present when its value at a location is higher than the mean value at all locations. The result is a table of values (called standardized residuals) for every combination of topographic variable and landscape group³.

The first part of the technique produces a set of contingency tables. Next, the frequency values in these tables, which reflect the frequency of the concurrence of topographic factors and landscape groups, were converted into normalized residuals using Haberman's method (Strahler, 1978) (See Table 1). These residuals reflect the importance of the independent variables in determining the dependent groups. The variables included:

IZDTM	elevations
IZSLOP	slopes (degrees)
AUTC	origin of the substratum,(alochthonous/autochthonous)
DRAIN	cumulative upslope area draining through a point
WATSAT	topographic index of water saturation
POTRAD	incoming short-wave radiation in winter (MJ/m ² /day)
REGLTH	regolith thickness (cm)
NOR	orientation toward northwest quadrant

(Note: The variable names correspond to image filenames.)

Table 1
Matrix in Q mode of the Haberman normalized residuals

	Groups									
Variables	1	2	3	4	5	6	7	8	9	10
POTRAD	-3.58	-2.14	1.3	1.09	0.43	0.43	-0.61	1.40	0.56	0.54
AUTC	3.12	1.84	-2.03	-0.03	-2.52	-4.01	-0.86	0.90	3.12	3.21
DRAIN	-2.66	0.37	0.21	1.61	-0.09	1.39	-0.88	-1.00	0.21	0.58

2. For a full discussion, see Strahler, A.H. 1978. "Binary Discriminant Analysis: A new Method for Investigating Species-Environment Relationships" *Ecology*, 59(1):108-116. In particular, the construction of the contingency table and the calculation of Haberman's residual value are found there.

3. These can be further analyzed using a Principal Components Analysis to determine how much each of the topographic variables contribute to the variation of landscapes across all groups. This analysis was not performed here.

	Groups									
IZDTM	-0.16	0.22	-1.34	-0.86	-2.06	-0.48	-1.74	1.58	1.67	4.65
NOR	1.63	1.31	-1.90	-1.05	0.89	-1.19	4.74	0.29	-1.58	-0.46
REGLTH	-1.71	-0.82	1.39	-0.18	0.67	2.16	-0.18	-1.81	-1.14	-0.24
IZSLOP	-0.7	1.85	-2.04	0.08	-1.45	-1.45	-1.70	3	2.39	1.97
WATSAT	-2.66	-1.04	1.38	0.78	-0.09	1.39	-0.05	-2.25	0.78	0.58

The table is interpreted as follows: each independent topographic variable (arranged in rows) contributes to a greater or lesser degree to the occurrence of the ten ecological landscape groups (arranged as columns). A high positive value in the table represents a high "attraction" between the independent variable and the corresponding group. A high negative value represents a high "repulsion" between the variable and the group. A low absolute value, of either sign, represents no meaningful relationship between the two.

A high value of either sign, suggests that the *presence* (a value higher than the mean) of a topographic variable explains greatly the occurrence of a landscape group. In the case of a high negative, the *absence* (a value lower than the mean) of that variable generally corresponds to the occurrence of that landscape type. For example, the absence of potential radiation (POTRAD) corresponds highly with landscape Group 1. This means that landscape Group 1 usually occurs under conditions of relatively low potential radiation.

Because the units of the residuals are in standard deviations, it is possible to take their absolute values and then arrange them hierarchically within each landscape group in decreasing order of association. For example, landscape Group 1 is explained in descending order by:

POTRAD	(-3.58)
AUTC	(+3.12)
DRAIN	(-2.66)
WATSAT	(-2.66)
REGLTH	(-1.71)
NOR	(+1.63)
IZSLOP	(-0.70)
IZDTM	(-0.16)

This means that landscape Group 1 seems most prevalent under conditions of relatively low potential radiation, with an autoctonous substratum, and little drainage. Slope and elevation do not figure significantly one way or another for this ecological group.

The significance of the results was tested. All variables tested significantly at the 95% level using the ANOVA except the values relating to cumulative area drainage (DRAIN) through a point and regolith thickness (REGLTH). It was possible that the threshold value was too high and the algorithm to calculate drainage partially unreliable. These variables proved significant enough to retain. AUTC was dropped altogether after the discriminant analysis because it did not prove statistically significant in determining ecological groups at all. All the other variables will be used for further analysis.

Having introduced the source of the sample data set, reviewed the statistical procedure employed, and derived the rela-

tionships between variables and landscape groups, we may employ these relationships to derive landscapes.

Producing GIS Images from Topographic Variables

Introducing the Procedure and Data Set

It is at this point that we begin the GIS analysis. We will apply GIS to test the degree to which the variables are consistent with the observed landscape. The goal is to produce maps of the potential distribution of each ecosystem group across the catchment as predicted by the BDA results and then compare them to actual sampled point observations.

We will do this for each landscape group by combining images of contributing variables to build an overall image of the predicted distribution of each group. The images of the independent topographic variables will be combined one at a time, starting with the most significant variable (as determined previously in Table 1). After each overlay combination, a chi-square test of agreement will be conducted with the predicted landscape against the landscape at actual sample points. At the point when the predicted landscape from an overlay of variables has reached good statistical agreement with the sampled data set, we stop overlaying the less significant variables and save the resulting image as a good approximation of the distribution of the landscape. We can repeat this process for all ten landscape groups.

Finally, combining the ten individual landscapes, we can examine the spatial distributions of a predicted composite landscape for the region.

The GIS layers containing the topographic variables correspond to the names in Table 1 (except AUTC which was removed from analysis after the BDA).

- a) Use IDRISI Explorer to view the contents of the data set and additional files processed for the exercise. Also, use DISPLAY Launcher to look at the files corresponding to the variable names in Table 1 (except AUTC which is not present in the database). The origin of the digital elevation model is a topographic map scale 1:1000 with 1 meter contour interval. Contours were digitized at 5 m intervals, rasterized to a 10 m pixel resolution, and interpolated. All other variables except regolith thickness were derived from the elevation model.⁴
- b) Display IZDTM with ORTHO. Do not use a drape image. Choose to change the minimum and maximum values for display. Set the minimum value to 2000 and the maximum value to 2500. Elevation values in the Izas catchment range between 2059 m to 2285 m. By setting a higher minimum value than the background value of 0, the contrast among heights is improved for display.

Note the data area is irregular. The total sample size of data, i.e., the number of pixels, is a subset of the complete set of pixels. The remaining pixels are background, typically having zeros as their values.

Binary Coding of Variables: Creating Boolean Images

Each of the topographic variables used in BDA here has a corresponding image in the data set. As seen above, these are continuous images showing the values of the variable across the study area. The slope variable (IZSLOP) for example, is shown in the image IZSLOP where each cell is given a value of slope in degrees. You will recall, however, that in the Binary Discriminant Analysis employed previously, topographic variables were defined (in Table 1) by the intensity of their positive or negative association with a given landscape group. The variables contributed to the landscape by their "presence" if they were above the mean value in a location and by their "absence" if they were below the mean.

4. The hydrological variables both were developed using TOPMODEL, a semi-distributed hydrological model that considers topography as a suitable indicator for gravitational potentials involved in responses based on surface flow (Beven & Kirby, 1979; Beven & Wood, 1983; Beven, 1987). Regolith Thickness was produced by kriging, using the values recorded in the sampling points. A linear semivariogram was assumed using a quadrant search (Davis, 1973).

To combine these variables based on the above statistically defined relationships, we must first break each continuous image of a topographic variable into two Boolean images. One will show only areas where the variable is above the mean, the other only showing areas below the mean. The mean values (shown in Table 2) were determined for each variable after normalizing their distributions using the Box-Cox transformation (Sokal, 1981). By reclassifying the images, the mean values of the variables can be used to divide each image into two spatial zones: those values equal or above the threshold and those below. We will create two Boolean images from each of our existing images. For each variable, one new image will show the distribution of pixels whose values are above the mean. The other will show the distribution of the below mean value pixels.

Table 2
Mean values of quantitative topographic variables

Variable	Mean	Upper Bound	Lower Bound
IZDTM	2165	2167	2164
IZSLOP	16.4	16.6	16.1
DRAIN	409.8	421.6	396.8
WATSAT	4.95	4.99	4.92
POTRAD	9.3	9.4	9.2
REGLTH	35	36	35

To build a landscape from these binary variables, we will combine them in the order of their significance to a landscape group. If a topographic variable is positively associated with a landscape, we will use the image of the above average topographic values. If it is negatively associated, we will use the image of below average values.

Using the mean values of Table 2, we reclassify each topographic variable to produce two Boolean images in this manner.

- c) Reclassify IZDTM by choosing the user-defined option in RECLASS to create an image called IZDTMB. Set all values less than 2165 to 0 and those greater than or equal to 2165 to 1. Look at the result in DISPLAY Launcher using the Qualitative palette. This image represents the distribution of the variable in the case of positive associations with an ecological group.
- d) To create the inverse image, called IZDTMB0, it is simpler to reclassify IZDTMB with Edit/ASSIGN in order to set values of 1 to 0 and 0 to 1, rather than employ RECLASS again. Open Edit and enter the following:

0 1

1 0

Save the file with the name BOOL and an .avl extension. Your data type is integer. Run ASSIGN, entering IZDTMB as the input image, BOOL as the attribute values file, and IZDTMB0 at the output file. View the result in DISPLAY Launcher .

We need to apply this process to the rest of the images. To minimize the burden of repetitious steps, one can use a macro file here.

- e) Run Edit and open the macro file called BOOLEAN. In this batch file, RECLASS and ASSIGN are used to create a new set of Boolean images from the input images. Each image is reclassified using its mean value listed in Table 2. For the output names, we used B for positive associations and B0 for negative ones, to produce the fol-

lowing:

IZDTMB	IZDTMB0
IZSLOPB	IZSLOPB0
WATSATB	WATSATB0
DRAINB	DRAINB0
POTRADB	POTRADB0
REGLTHB	REGLTHB0
NOR	NORB0

The NOR variable is already a Boolean image, so we need only make its inverse image and substitute NOR for NORB0.

- f) Go to the File menu and select Run Macro. Specify BOOLEAN as the file.

The resulting images should match the image names given above. These are Boolean images of each of the topographic variables which may be used in varying combination to produce predicted distribution images for each of the ten landscape groups.

Modeling of the Expected Distribution

The Modeling Procedure

Next, we model each ecological group's potential distribution. Various combinations of the Boolean images will give us this distribution. We first need to determine which variables are most significant for predicting the pattern for each group. We assume that the strength of each variable's influence on a group's formation is hierarchical as shown in the results of the BDA given in Table 1. We will successively combine topographic variables together using OVERLAY, one at a time, in the order of their strength of association.

At some point, each additional variable of lower rank that we add will describe less and less of the total distribution. We can estimate at which point this is reached. We use a chi-square test to compare the variables' spatial patterns to that of ground point data. These points have already been classified by ecological group. We calculate the chi-square value for each successive addition of a variable to the map of a landscape group distribution. When the chi-square value has reached a point of statistical significance, and begins to drop with the addition of any more variables, we stop. We have reached the point where combining the distributions of additional variables no longer helps describe the ecological group observed for the point data. Insignificant variables are dropped and the existing image is used to represent a final predicted spatial pattern for the landscape group. We can repeat this procedure for each landscape group.

Performing the Chi-square Tests

First, we must rank the variables for each ecological group by using the absolute value of the normalized residuals from the discriminant analysis (Table 1) and arranging them in decreasing order. For example, Group 9 variables fall accordingly:

IZSLOP
IZDTM
NOR
REGLTH
WATSAT
POTRAD
DRAIN

Recall that variable AUTC was dropped from the model (as mentioned earlier). Each of these variables has a positive or negative association with each landscape group. For a positive association, we use the Boolean image showing the distribution of values above the mean (our images with names ending in "B"). For a negative association, we use the Boolean image showing the distribution of values below the mean (our images with names ending in "B0"). In the case of Group 9, the images we test are:

IZSLOPB
IZDTMB
NORB0
REGLTHB0
WATSATB
POTRADB
DRAINB

A series of images is included that represent observed sample points for each ecological group called SAMP1-SAMP10. These will be used to test the accuracy of the results.

g) Use DISPLAY Launcher to view SAMP9 with the Qualitative palette. We can see that there are only 13 points.

In order to compare these points with Boolean images of the topographic variables, a simple chi-square test is inadequate. Because of the small sample size, we will use the chi-square corrected with the *Yates continuity correction*. We want to cross-tabulate these points which represent ground truth observations of a particular ecological group with data representing the expected distribution of the topographic variables. The degree of agreement between the two is one test of the predictive capacity of the topographic variables for an ecosystem.

By successively incorporating each variable in the order established by the BDA, we calculate successive chi-squares. We evaluate at each stage whether or not the combination of variables is sufficient to predict the distribution of a landscape group. In this way, it is possible to choose which variables are necessary for estimating the area of an ecological group.

To conduct this test, we will simply use the IDRISI modules OVERLAY and CROSSTAB. The module CROSSTAB produces cell counts of all possible combinations of values from two images and computes a simple chi-square calculation. However, this chi-square is not the one we need. We must calculate it ourselves to correspond to the sample points only.

The first topographic variable for Group 9 is IZSLOPB which we crosstabulate with sample point file SAMP9.

h) Use DISPLAY Launcher to examine IZSLOPB. Notice that there is no defined boundary to the study area. In

order to calculate the chi-square properly, we must take proper account of our values in the crosstabulation. We will need to give the variable area, the study area, and image background pixels all separate values.

1. *How would one create an image that shows all three categories: the distribution of the variable equal to 1, the other study area pixels equal to 2, and anything outside the study area equal to 3?*

The image STUDY is provided which represents the study area.

- i) Run OVERLAY and select the cover option. Specify IZSLOPB as the first image, STUDY as the second image, and call the output VAR1.
- j) Next, run CROSSTAB and specify VAR1 as the first image and SAMP9 as the second image. Choose Full cross-tabulation table as output and click OK.

The table which appears on the screen tells us the number of cells representing each combination of values from the two images. The rows represent our sample points (the observed values) and the columns represent our predicted distribution for Group 9, taking into account only slope as the predictor variable. The table should look like this:

	1	2	3	Total
0	1974	2512	2722	7208
1	8	5	0	13
Total	1982	2517	2722	7221

Column number three represents background pixels and is of no interest to us. We want to exclude it from our chi-square analysis. To compute the Yates continuity corrected chi-square, we are interested only in the relationship between columns 1 and 2 and rows 1 and 2. For use in the Yates formula, these variables can be read off of the table and are named as follows:

	1	2	3	Total
0	a	b	-	-
1	c	d	-	-
Total	e	f	-	-

These figures are substituted into the following formula to calculate Yates' corrected chi-square (X_c^2)

$$X_c^2 = (N (|a * d - b * c| - 1/2 N)^2) / (e * f * (a + b) * (c + d))$$

N represents the sum total of pixels in the study area: $N=e+f$

The numbered values a, b, c, and d represent portions of the predicted pixels agreeing or disagreeing with the sample set. The denominator represents the product of the sum of rows and columns. Note again that the third column of the contingency table refers to data outside the study area which are not of concern to us.

The formula tests the proportion of matching predictions against the size of the predicted set and gives a figure showing the significance of the number of correct predictions. The significance level for alpha = 0.05, according to a chi-square table, is 3.841. To be considered significant, the chi-square value must be at least equal to 3.841. Should the chi-square be equal to or higher than this number, we may assume that the landscape distribution predicted from the topographical variable(s) significantly mirrors the real-world distribution of that landscape type.

- k) Using the above formula, calculate the corrected chi-square from the CROSSTAB table of VAR1 (slope) and SAMP9. Determine whether it is above the significance level (3.841) for describing the distribution of Group 9 sample points.

2. *What is the chi-square value?*

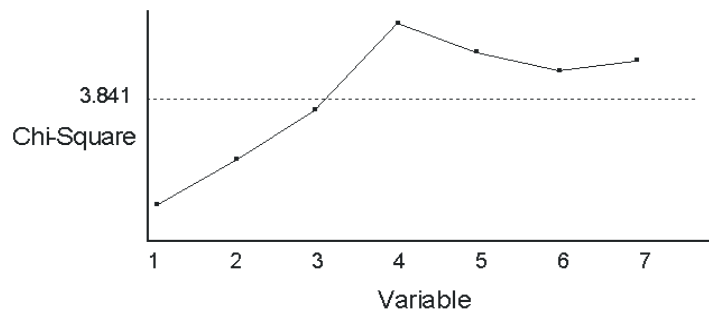
From the very low chi-square value, we see that predicting landscape 9 from slope alone is inadequate. We therefore want to incorporate the second variable IZDTMB (elevation) into the prediction of the Group 9 distribution. After using OVERLAY to incorporate this variable, we will calculate the chi-square again to test the significance of the new prediction.

- l) Produce an image containing the intersection of IZSLOPB and IZDTMB distributions. Use OVERLAY to multiply the two and call the output TEMP1.
- m) To test the significance of this distribution, first run OVERLAY to cover TEMP1 over STUDY. Call the output file VAR2. Next, CROSSTAB VAR2 with SAMP9, producing a cross-classification table only, and compute chi-square again.
- n) Repeat the process for NORB0 and REGLTHB0. Use OVERLAY to multiply TEMP1 with NORB0 to produce TEMP2. Multiply TEMP2 with REGLTHB0 to produce TEMP3. Cover the results on top of STUDY to produce VAR3 (from TEMP2) and VAR4 (from TEMP3). CROSSTAB VAR3 and VAR4 with SAMP9, recording again the number of observed values that match the predicted distribution reported in the table. Calculate chi-square again for each.

3. *What are the chi-square values for VAR2, VAR3, and VAR4?*

4. *What combination of variables is sufficient to significantly match the sample set?*

The chi-square values for successive generations of landscape group 9 tested against SAMP9 looks like this:



The maximum value of the chi-square corresponds to the combination of the first four variables. Each additional variable has declining significance to the potential distribution of Group 9. This means the additional variables reduce the model's predictive capacity instead of introducing new information. The combination image of the first four variables, which has the strongest association, represents the best expected distribution of Group 9. The image TEMP3 then corresponds to the optimal solution given the group of variables tested.

- o) Use IDRISI Explorer and rename TEMP3.rst as G9.rst. (Backup image GP9 is included in the data set. To use

this, rename GP9 to G9.)

In the first iteration, 8 out of the 13 sampling points assigned to this group fall within the patch of potential distribution. This patch extends over 49% of the total surface of the catchment. In the second iteration, a new condition is added, which reduces the patch to 32% of the surface, and leaves 7 sampling points inside it. However, that patch presents comparatively more sampling points than the former, and the chi-square increases as a consequence. The same situation appears again in the third and fourth iterations, when 5 observed points fall inside a patch which, fulfilling three conditions, covers 12% of the surface. On the contrary, with the fifth iteration, the additional condition reduces the patch with a dramatic relative loss of included pixels and the chi-square begins to decrease. This tendency continues up till none of the observed points fall within the combined area and the chi-square drops to zero.

- p) Go through the process of testing the variables again, this time for Group 8. Compare the points of SAMP8 to the variables listed in Table 1. Remember to use them in descending order. Compare the resulting chi-squares to the 95% (3.841) significance level.

5. *Which variables are used to simulate the distribution of Group 8?*

- q) Use IDRISI Explorer and rename the temporary image representing the group to G8 (backup image GP8 is included in the dataset).

We completed testing the remaining eight groups for you. The variables predicting each were selected and renamed into group files (all data are supplied with the exercise set in order to replicate this).

Discussion of Test Results

In those cases where the maximum is not above the 95% threshold value, we reject the potential distribution as credible whether or not these ecological groups are real -- they do not react sufficiently to the independent variables employed. This is the case with Group 2 and Group 4, reflecting that they either do not have morphological consistency or do not respond sufficiently to the variables employed. On the other hand, Group 10 shows the highest significance of all, and every variable was employed. While the distribution is acceptable by the standard used, introducing new variables probably would improve it.

Before proceeding, we need to change the format of the images. For further analysis, the areas showing a particular landscape group need to have the value of that group. In the images of landscape groups that we have provided, you will notice that this is the case. GROUP6, for example, gives a value of 6 to all the group's pixels.

- r) Use Edit/ASSIGN or RECLASS on images G9 and G8 to give each distribution its own group number. An old value of 1 in G9 will receive a new value of 9. An old value of 1 in G8 will receive a new value of 8 in the output image. Call the outputs GROUP8 and GROUP9.

Once the topographic variables best predicting the spatial distribution of each group are defined, it is possible to synthesize the evidence of topographic controls for the catchment as a whole through an examination of individual group responses. We can see where each landscape fits within the entire catchment.

Examination of Spatial Relationships Between Ecological Groups

After examining the development of groups under different topographic conditions, what can be deduced about the distribution of landscape groups among each other? What is their role in the catchment? To answer these questions, we need to examine the overlap between the potential distributions. For each pair of landscape group images, there is a possibility of overlap. Using CROSSTAB, we can calculate the percentage of cells that falls within an area of overlap, and use this spatial information to interpret relationships among ecosystems.

- s) CROSSTAB GROUP9 with GROUP10 to produce *both* an image and cross-classification table. Call the output

image G9_10. Count the number of cells that fall into the overlap category (where the values in both images are not zero). Record that figure in Table 3 below. Next, look at the result in DISPLAY Launcher. Repeat these steps for the remaining spaces in the upper half of the table. Give output names consistent with the above example.

Table 3

Overlap Between Groups (pixels)

	Groups						
	1	3	6	7	8	9	10
Groups :	(704)	(820)	(426)	(263)	(532)	(554)	(119)
1 (704)				90	137	66	0
3 (820)				0	0	0	0
6 (426)				0	0	0	0
7 (263)					0	0	0
8 (532)						327	0
9 (554)							

Overlapping among groups is not haphazard. There is a certain hierarchy in the development of ecosystems, such that some distributions are unique and specialized, while others are subsets included within a broader group. For example, Groups 6 and 7 have highly complementary distributions. Field study of the areas of overlap shows that 7 represents a specialized form of other landscape types, reflecting a higher hydric saturation. By using the maps of these areas of overlap for further field analysis and observation, it can be determined where one landscape type is subsumed as a specialized type of another landscape. This allows us to create a hierarchy which will aid in the creation of a final synthetic image.

Synthetic Modeling of the Landscape

In this last section, we view the expected distributions together and check for their spatial consistency with the topography. We also will create a composite image of the landscape groups by layering the individual images over one another.

To combine the 8 groups together, we will use OVERLAY. Considering that some groups are specialized forms of others, the order in which we combine them is essential. Some specialized landscape types must be absorbed in other, more general ones. By coupling field study with an examination of the overlap figures in Table 3, it is evident that Group 6 falls within Group 3 and Group 5, that Group 7 falls within Groups 5 and 1, and Group 10 falls within Group 9.

6. *Keeping in mind that some groups almost entirely overlap others, which option of OVERLAY will we use?*

In a few cases, some arbitrary decisions concerning overlapping priorities must be made. These overlays are numerous enough that a batch file might make it a little easier to conduct them.

- t) Create a macro file using Edit and call it MODEL. The following sequence of overlays will combine the various groups together into a composite image. Write the macro file to do the following:

```
OVERLAY    GROUP10    on    GROUP9    = COV1
```

GROUP8	on	COV1	= COV2
GROUP7	on	GROUP5	= COV3
GROUP6	on	GROUP3	= COV4
COV4	on	COV3	= COV5
COV5	on	COV2	= COV6
GROUP1	on	COV6	= ALL

When finished, save the file upon exit. To execute the macro, choose Run Macro from the File menu and specify MODEL as the file. If any errors occur, go back into Edit to see if there are any missing parameters.

- u) Use DISPLAY Launcher to look at the resulting image ALL with the Qualitative palette.

Notice large areas are unclassified. We have selected the most stringent conditions in our OVERLAY and chi-square test procedure and many pixels remain unallotted to any group. In the original study, it was concluded that some areas may be intermediate or initial stages of succession in a path leading to and from various groups. Hence, it is unnecessary to look for the highly specific topographic conditions where ecosystems produce their most typical structure. Instead, we may settle for less than optimal distributions. A more general, wider distribution would be more inclusive and would create a more fully covered image.

In practice, this means selecting those combinations of topographic variables that constitute the sub-optimum iteration immediately before the maximum similarity (high chi-square value) is reached. In creating the new distributions, the conditions can be relaxed by one variable for Groups 3, 5, 8, and 9. We used these assumptions to make new group images for these four landscape groups and created another composite image by combining them. This image is called GORC.

- v) For display purposes, the image GORC has a background value of 12 rather than 0. To compare ALL with GORC, use RECLASS on ALL, changing all old values between 0 and 1 to a new value of 12. Call the output image IZFINAL.
- w) Compare the final image, GORC, with IZFINAL by displaying both images with DISPLAY Launcher and placing the images next to each other

We can also compare the results in a three-dimensional view. This will help us to see the relationship of ecological group distributions with the topography.

- x) Run ORTHO to drape IZFINAL on IZDTM. Use a minimum value of 2000 and a maximum value of your choosing. Display with the Qualitative palette. Run ORTHO again with GORC as the drape image.

Conclusions

This exercise demonstrates the landscape approach in mountain ecosystem modeling. In this case, we briefly demonstrated deriving new information about the relationships of ecosystem groups to topography and among one other. A spatial component can in this way be incorporated into ecosystem identification by using GIS along with other statistical and classification packages.

While topographic factors were shown to be important determinants in ecological dynamics, obviously there is a need to understand these relationships relative to scale. Spatial patterns, in part, are dependent upon the scale of observation, as are the relationships interpreted from those patterns. At a different scale, the relationships among organisms become more significant than those relating to topographic variables. At smaller geographic scales (covering much larger areas), the significance levels of topographic variables to predict spatial patterns of ecosystems might increase.

It is also crucial to keep in mind that most landscapes have been carved not only by the powerful determining influences of topography but by human actions as well. Human presence in a landscape, traced through complex historical patterns of use, may contribute greatly to landscape formation, reconstruction, and maintenance. This fact requires some consideration as the technique is brought to bear in other regions. The spatial component of this analysis does, however, allow us to observe and predict how environmental systems are distributed at specific scales. It also allows us to consider the processes that link landscapes to larger phenomena. These principles of analysis may be useful whether the processes being considered are physical, biotic, or human.

References

- Beven, K.J. (1987) "Towards the use of catchment geomorphology in flood frequency predictions." *Earth Surface Processes and Landforms*, 12: 69-82.
- Beven, K.J. and Kirby, M.J. (1979) "A physically based variable contributing area model." *Hydrological Sciences Bulletin*, 24(1): 43-69.
- Beven, K.J. and Wood, E.F. (1983) "Catchment Geomorphology and the dynamics of runoff contributing areas." *Journal of Hydrology*, 65: 139-158.
- FAO (1985) *FAO-UNESCO Soil Map of the World*, 1:5000, revised legend. FAO, Rome.
- Hill, M.O. (1979) *TWINSPAN - A FORTRAN Program for Arranging Multivariate Data in an Ordered Two-way Table by Classification of the Individuals and Attributes*, Section of Ecology and Systematics, Cornell University, Ithaca.
- Hill, M.O., Bunce, R.G.H. and Shaw, M.W. (1975) "Indicator species analysis, a divisive polythetic method of classification, and its application to a survey of native pinewoods in Scotland." *Journal of Ecology*, 63: 597-613.
- Jackson, D.A., Somers, K.M. and Harvey, H.H. (1989) "Similarity coefficients: measures of co-occurrence and association or simply measures of occurrence?" *The American Naturalist*, 133(3): 436-453.
- Podani, J. (1988) SYN-TAX III, User's manual. *Abstracta Botanica*, 12, Suppl.1: 1-183.
- Sokal, R.R. and Rohlf, J.R. (1981) *Biometry*. W.H. Freeman and Company, New York.
- Sneath, P.H.A. and Sokal, R.R. (1973) *Numerical Taxonomy*. W.H. Freeman and Company, San Francisco.
- Strahler, A.H. (1978) "Binary Discriminant Analysis: A New Method for Investigating Species-Environment Relationships" *Ecology*, 59(1):108-116.

Answers

Exercise 1

1. To determine the rows and columns in the image, it is important to know the total length of the image in X and Y dimensions. Subtracting the minimum X from the maximum X coordinate will produce the total length of columns in the image. In the same way, subtracting the minimum Y from the maximum Y coordinate produces the total length of rows in the image. To determine the specific rows and columns matching the grid system of the raster image, we divide each of these measures by the corresponding pixel resolution. For example, to determine the total number of columns in an image, we divide the total length of columns (Xmax-Xmin) by the pixel size (in the X dimension). The total number of rows is also determined by dividing the total length of rows (Ymax-Ymin) by the length of the corresponding pixel dimension. In many cases, the spatial resolution (pixel size) of images is the same (e.g., 30 by 30) so that the X and Y dimensions of a pixel will be the same.
2. For many places, the change in elevation is greater than 250 per 90 meter horizontal change.
3. The mean filter had the effect of averaging or smoothing the values relative to one another. This is because it recalculates new pixel values based on the mean of each set of 9 (3 by 3) pixels. See the discussion of FILTER in the IDRISI Manual.
4. The difference in area figures for slopes greater than 45 degrees is caused by changes in spatial resolution (pixel size) of the various images. For example, a more detailed slope can be produced from a DTM image with a spatial resolution of 7 meters than from a DTM image with a 90-meter spatial resolution. Since the area calculation does not involve the whole map but only the portion with slopes greater than 45 degrees, the more detailed slope maps will have larger portions devoted to slopes greater than 45 degrees.

Exercise 2

1. Plantations are a new landuse category that did not exist in 1972.

Table 1. Landuse in 1972 and 1989

Landuse type	Area, 1972 (ha)	Area, 1972 (%)	Area, 1989 (ha)	Area, 1989 (%)	1972-1989 Change (%)
Bari (terraces)	388.7	31	505.4	40	9
Khet	82.9	6.5	106.1	8.4	1.9
Forest	248.8	19.6	259.1	20.4	0.8
Plantation	0	-	121.1	9.6	9.6
Grassland	255.3	20	127.6	10.1	-9.9
Shrubland	277.2	21.9	120.7	9.5	-12.4
Others	14.6	1.2	27.6	2.2	1.1
Total	1267.5	100	1267.6	100	

2. Forest cover (plantations) and Bari agriculture have expanded each 9% of their original extent in 1972; grassland and shrubland have decreased significantly. The total area of forest/plantation has grown. This does not necessarily mean that no deforestation has taken place in particular areas but does show a net gain in tree cover over the entire area. Irrigated agriculture (Khet) did not expand significantly because water availability has limited its growth.

Table 2. Landuse dynamics, 1972-1989

(Significant landuse changes shown in bold and percentage calculated)

Legend Number	Landuse, 1972	Landuse, 1989	Area, ha	% Change, 1972-89
2	Bari	Bari	321.5	
3	Khet	Bari	4.3	5.2
4	Forest	Bari	36.7	14.7
5	Grassland	Bari	69.2	27.1
6	Shrubland	Bari	73.6	26.6
7	Others	Bari	0.1	
8	Bari	Khet	11	2.8
9	Khet	Khet	66.2	
10	Forest	Khet	7.8	3.1
11	Grassland	Khet	7	2.7
12	Shrubland	Khet	11.9	4.3
13	Others	Khet	2.1	
14	Bari	Forest	25.8	6.6
15	Khet	Forest	4.4	
16	Forest	Forest	130.6	
17	Grassland	Forest	16	6.3
18	Shrubland	Forest	81.4	29.4
19	Others	Forest	0.7	
20	Bari	Plantation	5	1.3
21	Khet	Plantation	0.1	
22	Forest	Plantation	55	22.1
23	Grassland	Plantation	36.7	14.4
24	Shrubland	Plantation	24.7	8.9
25	Others	Plantation	0.1	
26	Bari	Grassland	11	2.8
27	Khet	Grassland	1.3	

28	Forest	Grassland	3.6	
29	Grassland	Grassland	83.3	
30	Shrubland	Grassland	29	10.5
31	Others	Grassland	0.3	
32	Bari	Shrubland	10	2.6
33	Khet	Shrubland	2.5	
34	Forest	Shrubland	15	6
35	Grassland	Shrubland	38.4	15
36	Shrubland	Shrubland	54.5	
37	Others	Shrubland	0.4	

3. Category 1 represents background areas, which are the same for 1972 and 1989.

4. Bari agriculture and forest cover have expanded at the expense of grassland and shrubland. The fact that almost 22% of the 1972 forest cover were under forest plantations in 1989 means that these forested areas have been put under the plantation project. They may or may not have undergone a loss of tree cover in the intervening period to justify their change in status. With the available data, we can only assess the overall trend in change but cannot explore the internal land use dynamics during the period of 18 years.

5. The elevations range is approximately between 900 m and 1800 m. The lowest part is in the northeast, the highest in the south, west and northwest “corners” of the sub-watershed. A histogram of the image values may be employed to examine this range.

6. Minimum elevation of the model is around 920 m. Contour interval of the map was around 50 m.

7. We chose 900 as the upper limit of the first class because we wanted to make sure that it was lower than the minimum elevation in the model, so that none of the cells from the elevation model were assigned the value of 0. We could have it set to any number between 0 and 900. If we had chosen 1000 m as the upper limit, part of the elevation model would have fallen into the range and would have been incorrectly classified.

8. 1200 – 1400 m is the predominant elevation range in the area.

9. The 800 – 1000 m elevation class is missing – meaning that there was no Bari agriculture expansion at these elevations between 1972 and 1989.

Table 3. Landuse dynamics in relation to elevation

Elevation Class	Bari Expansion		Forest Expansion		Grassland Loss		Shrubland Loss	
	ha	%	ha	%	ha	%	ha	%
800-1000 m	0	1	8.1	6.7	0	0	0.4	0.2
1000-1200 m	53.7	29.2	65.3	53.9	13.6	8.1	104	47.3
1200-1400 m	80.8	43.9	14.4	11.9	68.5	40.9	97.6	44.4
1400-1600 m	40	20.9	32	26.4	75.8	45.2	16.5	7.5
1600-1800 m	9.4	5.1	1.4	1.2	9.6	5.7	1.5	0.7
Total	183.9	100.1	121.2	100.1	167.5	99.9	220	100.1

10. Most of the agricultural expansion is occurring upslope, in areas above 1200 meters. Grazing land loss is occurring upslope, more than half of it in critical erosion risk areas above 1400 meters. Shrubland loss is occurring to a great degree in the middle 1000 to 1400 meter range. Agriculture has expanded into 130 ha of higher elevation areas versus only 47 ha of plantation expansion in this range.

Table 4. Landuse dynamics in relation to slope

Slope class	Bari Expansion		Forest Expansion		Grassland Loss		Shrubland Loss	
	ha	%	ha	%	ha	%	ha	%
0 – 5%	1.1	0.6	0.3	0.2	0.5	0.3	1.2	0.6
5 – 20%	5.8	3.2	5	4.1	3.1	1.9	9.7	4.4
20 – 35%	22.4	12.2	20.1	16.6	15.8	9.4	34.2	15.6
35 – 50%	55.5	30.2	37.9	31.3	35	20.9	72.9	33.2
> 50%	99.1	53.9	57.8	47.8	113.1	67.5	102	46.4
Total	183.9	100.1	121.1	100	167.5	100	220	100.2

11. 54% of agricultural expansion is occurring in slopes higher than 50% while 52% of forest plantation is occurring in slopes lower than 50%. Grass cover is declining in middle to high slope areas. Agriculture is being pushed into high elevation and slope regimes displacing soil-holding grass cover. Forest plantations, aimed at halting erosion, are being established in lower risk, low slope, and elevation areas. This means that the slope stabilization component of the afforestation program has not been adequately addressed, and the highest and most steep slopes will be at high risk of being eroded under marginalized agriculture.

12. The majority of the slopes face east and west but there are more northern slopes than southern.

Exercise 3

1. The elevation ranges from 315 m to 1905 m, a striking variation, though not unusual in this part of the Nepalese Himalaya.
2. The study area is defined in the LANDUSE file by 402 columns by 455 rows. At a cell resolution of 60 m per pixel, this comes to 24.12 km by 27.30 km. The total coverage of the study area, then, is around 658.5 square km.
3. The settlements and roads, not surprisingly, fall in and along river valley bottoms.
4. The value range of SLPFR is between 0.99 and 114.41. The low friction values fall along the flat valley bottoms while the prohibitively high values fall along valley walls, especially in parts of the north and southwest.
5. Each pixel is 60 meters wide. The buffer should include all pixels less than 120 meters away. This will cover a single pixel width but no more.
6. 15 villages fall within the range of center 2. Only 5 fall within the range of center 9. Keep in mind this is determined based on the range as measured from the centers to the villages.
7. Center 9 is closest to village 50 based on the measurement from the village. Measuring in the other direction (from the centers), village 50 is closest to center 6. This can be seen in the values extracted from the allocated file COSTRGN.

Exercise 4

1. The aspect of an image refers to the exposure of slopes measured in degrees from 0 to 360. Aspect is thus based on a change in elevation so that in areas that are flat with no slope, negative values are produced to represent the nature of plane exposure.
2. The maximum value in the image is 1.37, which is about 0.606 less than the Solar Constant of 1.98 ly/min/cm².
3. The minimum data value is 1.17.
4. High temperature concentrations are found in the northwest moderate elevation areas.

Exercise 5

1. The minimum and maximum values are:

Image	Lower bound	Upper bound
NEPC2	5	42
NEPC3	1	54
NEPC4	13	70

2. In each of the three images, most of the values were bunched in a constricted value range. These would all take similar colors or gray shades during display. By spreading the values across a wider range, the variations in the data become visibly clearer.
3. The values of NRATIO fall between -1 and +1. This is a natural artifact of a normalized ratio with identical input value ranges.
4. The ratio provides a normalized value for each pixel, which highlights variation for even shadowed areas. Consider pixels with low values in both the Red and Infrared bands. While both bands have experienced dampening, they have experienced it similarly, and the proportion of one value to the other remains roughly the same. The ratio image shows these proportions across the area, showing relative values between bands. Pixels with low and hard to interpret absolute values

in the input images may take clearer and more meaningful values when viewed as a proportion. Here, values in COMP1, otherwise lost in the shadows, can be seen with distinguishable values in COMP2. This is especially true of the deeply shadowed areas on the southern slopes of the steep valleys.

5. In MAP3, the coverage of hay/pasture is larger as is developed. More importantly, in MAP3, we can distinguish a large mixed wood category in the mountainous western region. This is lost in the clustered MAP1 image. It is likely that this is an artifact of the shadowing in the area. The topographic effect is uncorrected in MAP1, and information may have been lost.

Exercise 6

1. A base image must first be created which gives values of 2 and 3 to pixels within and outside the study area respectively. This base image could be used in combination with each of the Boolean images. By overlaying and covering the base image, each Boolean image would give a new value of 1 to all pixels within the distribution of the given variable. The pixels outside the distribution would retain a value of 2, and the background a value of 3.

2. The Yates corrected chi-square value for the crosstabulation of VAR1 and SAMP9 is 0.418. This is well below the level suggesting significant agreement. The distribution of slope alone does not therefore explain the distribution of landscape Group 9.

3. The corrected chi-square values for the cross tabulation of VAR2, VAR3, and VAR4 with SAMP9 are 1.917, 2.798, and 6.005 respectively. Only the last of these surpasses the threshold of significant agreement. Only the distribution of the combination of the first four variables is therefore sufficient to explain the distribution of Landscape Group 9.

4. The χ^2_c for the combination of variables 1, 2, 3, and 4 passes the significance mark. Therefore, slope, elevation, north-west orientation, and thickness of the regolith are sufficient to explain the distribution of Landscape Group 9.

5. The combination of the first three variables (high slope, low water saturation, and a thick regolith) has a chi-square of agreement with SAMP8 surpassing the minimum level of significance. The chi-square value of a combination of the first four variables actually reduces the chi-square below the level of significance. Therefore, the image showing distribution of the combination of these first three variables makes a good map image for the distribution of Landscape Group 8.

6. The cover operation is needed here to overlay one group over another so that the first “swallows” the second where they overlap. Any other operation (adding, multiplying, dividing, etc.) would result in the creation of a new value in the overlap area.