GIS-based Modelling and Ecology: A Review of Tools and Methods

Ioannis N. Vogiatzakis

Geographical Paper No. 170

September 2003

The author is a Lecturer in the Department of Geography, The University of Reading, Whiteknights, PO Box 227, Reading RG6 6AB
Telephone: 0118 378 8733
Email: i.n.vogiatzakis@reading.ac.uk

Series Editor: A M Mannion
a.m.mannion@reading.ac.uk
Abstract

The rapid development of computers and associated software during the last thirty years has led to the expansion (emergence) of Geographical Information Systems (GIS). Geographical information systems (GIS) are computer-based systems designed specifically to facilitate the digital storage, retrieval, and analysis of spatially-referenced environmental data. Coupled with ecological modelling, GIS can provide significantly increased opportunities for detailed environmental resource inventory and analysis and show considerate promise for extensive use in nature conservation. The paper introduces these two concepts and discusses the role of GIS-based modelling in nature conservation focusing on the predictive models for species occurrence, plant community occurrence and habitat suitability. The importance of Digital Elevation Models and their derived properties in these ecological studies is explained. Emphasis is placed on empirical or inductive modelling based on field observations. The generic steps of empirical modelling are described and demonstrated by a case study in Lefka Ori, Crete, Greece. Tools such as fuzzy mapping and geostatistics have a potential role to play in improving the level of information and therefore in the understanding of species and plant community distribution.

Keywords: Crete, DEM, empirical modelling, habitat, predictive mapping
1. Introduction

Making observations and recording information about the earth for visual representation has been a major human activity since the appearance of the first civilisations (Holt-Jensen, 1999). Maps have always been invaluable tools in ecological studies, providing spatial as well as attribute information, e.g. maps showing species distribution, extent and distribution of reserves, or the distribution of vegetation communities (Wadsworth and Treweek, 1999). The complexity of ecological problems, however, requires a diversity of information from a range of sources as well as analytical techniques that cannot be applied to conventional maps. Until recently the traditional means used for storing spatial information, i.e. the paper map, had to deal with storage limitations and restricted capabilities for updating and analysing spatial data. Those limitations have only recently been overcome with the evolution of computer-aided cartography. Computers have provided the capacity and the speed for the analysis of large and complex databases (Bernhardsen, 2002). Moreover, parallel developments in spatial data processing disciplines, such as topography, photogrammetry, remote sensing and geography, provided the opportunity for different sets of spatial data as well as techniques to be linked together, leading to the final “shaping” of Geographical Information Systems (GIS) (Burrough and McDonnell, 1998).

The first GIS was developed in Canada in 1960 in order to solve problems related to data handling and storing. However, during the seventies and eighties the use of GIS expanded due to increased governmental involvement in natural resources development and increased pressure on those resources. The opening of a large international market for this technology benefited by developments in the IT industry (e.g. low cost, high-speed processors), resulting in the adoption of these systems by a broader range of users. Nowadays many developing countries are employing GIS for applications in cadastral mapping and resource analysis (Burrough and McDonnell, 1998).

Although many definitions of GIS can be found in the literature, Burrough (1986, page 6) describes it as “a powerful set of tools for collecting, storing, retrieving at will, transforming and displaying spatial data from the real world for a particular set of purposes”. A GIS comprises three major components: computer hardware, computer software and organisational context (Burrough and McDonnell, 1998). The set of software operating within the computer
environment allow a wide range of analyses to be performed covering three main functions; the input of spatial data, the transformation of that data and the output of spatial and associated statistical data. A common feature of most GISs is that data may be derived from many sources and are recorded at different levels of resolution (Elston and Buckland, 1993). Data are usually stored in databases which are comprehensive collections of related data in logical files that are collectively processed using tabular form (Bernhardsen, 2002). The advent of GIS, provided ecologists with a tool for analysing the influence of the environment on ecological processes (Johnston, 1998). Consequently traditional cartographic techniques were gradually replaced.

Ecological modelling involves another well-established group of techniques also used in environmental studies (Jørgensen, 1994). The use of a model in order to express a physical process is at best a simplification of real world processes (Steyaert, 1993), since these processes are very complex and operate at a variety of scales in space and time. Nevertheless, at its best, modelling may enhance understanding of ecological processes. Ecological models are techniques that simulate ecological systems and processes. Ecological modelling combines mathematical modelling, systems analysis and computer techniques with the ecology and management of the environment and its natural resources (Jørgensen, 2003). The two main objectives of ecological modelling are explanation and prediction. When the model characterises the direct interactions of system components to understand the processes involved, it is termed descriptive. On the other hand, when it examines the system response to a potential change of underlying factors or their interactions (e.g. climate change), it is termed prescriptive (Berry, 1995).

Ecological data sets have two distinct characteristics if compared to other kinds of data: they are multivariate and location specific. Although historically ecological modellers have focused on changes in time at single sites or small geographical areas, during the past two decades, they have started to incorporate spatial pattern in the models and apply them in large geographic areas (Hunsaker et al., 1993). Most GISs lack the predictive capabilities to examine complex problems, whereas numerically oriented models lack flexible spatial analytic components to respond to the spatial character of ecological problems (Parks, 1993). The fact that both methods shared a common goal, i.e. addressing ecological problems, but at the same time exhibit limitations when
employed separately for that purpose, has led to their integration (see reviews in Goodchild et al. 1993 & 1996; Skidmore 2001).

Ecological modelling with GIS involves their complementary use for addressing ecological problems. There are two ways of linking ecological models with GIS. The first is to run the model outside the GIS and use the GIS for pre-processing (e.g. co-ordinates transformation, projection change) or post processing of the data (e.g. cartographic and visual display, simple spatial analysis). For example, Acevedo et al. (1996) presented a forest dynamics model linked to a GIS which was employed as a post-processing tool for display and analysis of model results. Generally speaking however, this linkage can be problematic due to the lack of common data models, structures and common interface.

Alternatively the GIS and modelling module can share the same data structures allowing systems to interact with the same database. In this case models are calibrated and run directly in the GIS, using the GIS command language (Goodchild, 1993). An example is given in Fedra (1996), where a modelling framework designed for assessing global change scenarios uses spatial information processed by an external GIS such as Arc/Info or Grass.

The advantages resulting from the fully integrated approach are (Parks, 1993):

- input variables are defined as continuous surfaces, thus areas different from the average can be recognised in the model;
- spatially dependent operators such as effective distance can be included;
- ability to deal with error propagation in the model.

GIS and ecological modelling have been employed in studies of terrestrial, freshwater, and marine ecosystems. Some examples include predicting forest composition and structure (Ohman et al. 2002) mapping benthic habitats (Urbanski and Szymelfenig 2003) and analysis of nutrient loads in rivers and streams (Pieterse et al. 2003)

In terrestrial ecosystems, the integration of GIS and modelling owes its spatial dimension to the discipline of landscape ecology. In contrast to traditional ecology, which has not focused on
recording information on the geographic location of species (Wadsworth and Treweek, 1999), landscape ecology is interested in the effect of location and spatial interactions in ecology. These interactions, such as the distribution of species in relation to other species or to the physical environment, which had been previously overlooked or oversimplified, gained importance with the adoption of GIS (Johnston, 1998). Although primarily a tool used in landscape ecology (Johnston, 1998; Haines-Young et al., 1993), GIS is now used for a wide range of applications for answering questions on the ecology and distribution of individual species and communities (Scott et al. 2002)

Inductive or deductive modelling approaches can be employed. Inductive or empirical approaches are based on the analysis of field collected data. Thus prediction is induced from empirical observations. Deductive or theoretical approaches are based on accepted theories on relationships between phenomena (Berry, 1995).

The aim of this paper is to review the most common tools and methods currently used for ecological modelling with emphasis placed on empirical modelling and its applications for vegetation and habitat mapping
2. Nature conservation and GIS

As natural habitats worldwide are being destroyed or converted to other uses, species supported by those habitats are inevitably threatened (Groombridge, 2002). The first step towards ensuring the long term persistence of the elements that comprise biodiversity is to develop the basic information required for their effective management (Jennings, 1995). This information should incorporate data on the distribution, biology and habitat requirements of species in danger. Where those data are available, GIS provides a means of rapidly reviewing the distribution and conservation status of several components of biodiversity. This information can then be used for detailed resource inventory and for decision-making purposes in nature conservation and management (Scott et al., 2002). However, even where information is poor or non-existent, GIS techniques can be used to predict species distribution patterns based on limited field data (Austin, 1998).

Some of the most common GIS applications in nature conservation include identifying and setting priorities either for further action and research (Kiester et al., 1996) or in the context of environmental impact assessment prior to development projects (Bojorquez-Tapia et al., 1995). Moreover, the integration of GIS together with other quantitative techniques for mapping/modelling vegetation communities (Brzeziecki et al., 1995), as well as habitat suitability (Guisan and Zimmermann, 2000) is a promising aspect of this relatively new field.

Habitat Modelling

The term “habitat” has been used in many ways in ecological studies. According to Spellerberg (1992), habitat can be defined as “the locality or area used by a population of organisms and the place where they live”. The principles applied to habitat modelling are analogous to the ones for predictive vegetation mapping (section 2.4). Where census is difficult, inferring possible distributions can be an alternative solution in species-habitat studies. Habitat factors should be considered in these studies, since species ranges and richness are often correlated with these factors. Therefore, prediction is possible for areas where reliable maps do not exist. The habitat features assumed to influence species distribution patterns are mapped and subsequently analysed within a GIS. Literature or empirical data can supply information on species-habitat relationships.
that can then facilitate the identification of habitat suitability and consequently the creation of potential distribution maps (Wadsworth and Treweek, 1999).

Since vegetation provides food and shelter to wildlife, it can be used as a surrogate for other habitat factors in the modelling process. However, several factors complicate the use of vegetation, imposing limitations to a habitat model. For example, the structure of vegetation rather than its floristic composition can be important to some taxonomic groups such as birds. Species have different habitat requirements in different parts of their range and at different times in their life cycle (Scott et al., 2002). Therefore, the presence of a vegetation type within a species’ range does not necessarily mean presence of the species under investigation. When the end product of habitat modelling is a predictive habitat suitability map, then no distinction should be made between predictive vegetation mapping and habitat modelling (Franklin, 1995).

Predictive models using the habitat-association approach have been employed in ecology to estimate species population sizes, geographical ranges as well as identifying potential impacts of habitat change (Fielding and Bell, 1997; Schwartz et al., 2001). Sperduto and Congalton (1996) used GIS to locate potential habitat for a rare orchid, *Isotropia medeloides*. Two scenarios were employed: one based on a simple overlay model and another on a weighting scheme after using a chi-square test of the significance of each parameter taken into account. Potential habitat maps were produced and field evaluated to assess the accuracy of the predictive models. The chi-square model was found to be more accurate (78% accuracy) than the equal weight model (57% accuracy).

In a study of nest site selection of two hawk species in a managed forest in Central Georgia, USA, Moorman and Chapman (1996) used GIS integrated with non-parametric statistical analysis. The study involved a macro-habitat and a microhabitat approach. In the first case, GIS was employed for habitat-type classification, while in the second, the important parameters related to habitat selection were determined based on distance operations. The potential relationship between habitat selection and nest success was then discussed. A good example demonstrating the GIS potential in natural resource management is the study by Brown et al. (1994) on the conflict of forestry practice and caribou habitat in Mount Revelstoke and Glacier
National Parks, Canada. First, three seasonal models were used to assess habitat suitability for caribou in the area. The maps were then compared with the harvesting plan within the forest, providing information on the location and extent of the conflicts. Management scenarios were finally proposed for wildlife protection with a minimum impact on timber yield.

Wu and Smeins (2000) employed GIS for predicting the occurrence of eight rare plant species in southern Texas, USA. They developed regional, landscape and site scale models based on a range of sources literature information, field visits and existing maps. They concluded that the suitability of a model is application dependant. In other words although regional and landscape model provide a good and fast overview for broad conservation assessment, site based models are more accurate but more expensive to use due to their data requirement needs. Finally, Gurnell et al. (2002) developed a GIS-based habitat suitability map for Sciurus vulgaris (red squirrel) in an area of 2800ha of Thetford forest, East England. The study aim was to assess the effects of forest management on the species population. Based on different habitat patches linkage as well as different minimum area patch sizes they explored 120 different scenarios and predicted that an increase of 82% of habitat suitability within the forest by 2015 as a result of existing and future management practices.
3. Predictive Vegetation Mapping

The most accurate portrayal of vegetation communities and the information related to them can be obtained by mapping. The field of vegetation mapping has resulted as “a fruit from the union of botany and geography” (Küchler and Zonneveld, 1988). The procedure initially involves the determination of the vegetation units using a classification scheme and then mapping the spatial extent of these units over the study area. Vegetation patterns are determined by environmental factors that exhibit heterogeneity over space and time, such as climate, topography, soil, as well as human disturbance (Alexander and Millington 2000). The need to map these patterns over large areas for resource conservation planning and to predict the effects of environmental change on vegetation distributions has led to the rapid development of predictive vegetation mapping.

Franklin (1995) defines predictive vegetation mapping as predicting the vegetation composition across the landscape from mapped environmental variables. The first attempt at predictive vegetation mapping was the work of Kessell in the late 1970s (quoted in Franklin, 1995). He called his approach “gradient modelling”, a term that was adopted by Austin (1987) and refers to the way species distributions respond to environmental gradients. Predictive vegetation studies start with the establishment of a model between the vegetation units and the mapped physical data, followed by the application of that model to a geographic database and over a wide range of spatial scales (Figure 1). With the advent of GIS and remote sensing techniques as well as the availability of digital maps of environmental variables such as topography, geology and soils, the development and implementation of predictive models has found a wide range of applications in vegetation studies (see for example Millington et al., 2002).

A number of modelling methods (Table 1) are available and can be classified into three main types (van Etten, 1998):

- heuristic methods based on a combination of field data and expert knowledge are used to define the important environmental parameters for the vegetation;
- decision trees attempt to define boundaries in environmental space for different vegetation communities, based on the middle value of a given predictor variable;
statistical models employ mainly regression to predict the value of the response variable if continuous, or the probability of class membership of a variable if categorical.

<table>
<thead>
<tr>
<th>Study</th>
<th>Modelling Method</th>
<th>Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Van Etten, 1998</td>
<td>Heuristic</td>
<td>Communities</td>
</tr>
<tr>
<td></td>
<td>Decision trees</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Generalised linear models</td>
<td></td>
</tr>
<tr>
<td>Zimmermann et al., 1999</td>
<td>Logistic regression</td>
<td>Species and communities</td>
</tr>
<tr>
<td>Wu and Smeins (2000)</td>
<td>Experts knowledge</td>
<td>Species</td>
</tr>
<tr>
<td>Carmel et al. (2001)</td>
<td>Logistic and Linear Regression</td>
<td>Communities</td>
</tr>
<tr>
<td>Miller and Franklin (2002)</td>
<td>Generalised Linear Models</td>
<td>Alliances</td>
</tr>
<tr>
<td>Vogiatzakis et al (in press)</td>
<td>Logistic Regression</td>
<td>Community types</td>
</tr>
</tbody>
</table>
Figure 1 Predictive Vegetation Mapping Stages
In particular regression analysis techniques have been traditionally employed in ecology for determining the relationships between species/communities and environments based on observations at given sites (Jongman et al., 1995). The advances in Generalised linear models (GLM) and generalised additive models (GAM) have led to their extensive application in ecological research as demonstrated by an increasing number of published papers (see review in Guisan et al. 2002).

Following the community-continuum debate in plant ecology (Crawley 1997) two main approaches in predictive mapping have been employed; the species-based approach and the community-based approach. A refined development of the continuum theory is the application of Generalised Linear Models to the prediction of species distributions (Guisan et al., 2002). Modelling the distribution of individual species is favoured from a theoretical point of view. Since the species is the basic ecological unit, predictive mapping of species distributions presents fewer definitional uncertainties or abstractions than communities (Wildi, 1998). The community approach has been criticised mainly due to the vague notion of the community concept and the fact that the classification needed in this approach is method-dependant (Wildi, 1998). Moreover, it has been argued that the approach is employed as an alternative in cases where sufficient data to model species distributions are lacking (Franklin, 1995). However, it is from the practical point of view that community modelling is preferable to species modelling. The main reason being that it is difficult to fully integrate the species ‘individualistic’ characteristics into static equilibrium models (Zimmermann and Kienast, 1999). Moreover, the observed presence of a species on a specific site is always an expression of its realized niche, which is defined not only by environmental but also by biotic factors and the competition with other species present (Tivy, 1993).

The results of both species and community modelling approaches vary according to the variables used and the data available. For instance, rare species have been modelled with success where adequate information was available (Sperduto and Congalton, 1996). In other cases, however, species models gave accurate results for common but not for rare species under investigation, since the right variables were not chosen (Cherill et al., 1995).
Studies employing the community modelling approach have also been successful. Zimmermann and Kienast (1999) used logistic regression and developed predictive models for species and communities in alpine grasslands of Switzerland. Communities were found to reflect small-scale topographic differences in vegetation; thus they were predicted with higher accuracy than species. This was because species-realised niches are more difficult to interpret in terms of (DEM) topography, since they are complex. The study concluded that community patterns are easier to simulate because of their relatively uniform response to environmental gradients.

Studies employing ordination and classification techniques used in vegetation analysis have also been used with GIS, although they are relatively few. For example, Fairbanks et al (1996) employed Canonical Correspondence Analysis (CCA) with a GIS database in order to study the changes in the floristic composition of chaparral and pine forest communities in California under a climate change scenario. Another case study in the Delta region in Haringvliet, S.W. Netherlands, used TWINSpan and Detrended Correspondence Analysis (DCA) to study the environmental factors that determine the zonation of vegetation types present (van de Rijt et al., 1996). Logistic regression was employed to describe the probability of occurrence of the vegetation types in relation to the frequency of flooding under grazed and ungrazed conditions.

Comparisons between different types of models can also be found in the literature (Moisen and Frescino 2002). For example, van Etten (1998) used three types of models to predict the distribution of the plant communities in the Hamersley ranges in Western Australia, namely a conceptual, a decision tree and a statistical model. The study concluded that although the models made similar predictions, differences occurred among the cover values of the maps obtained from each method.

A recent case study on the Mojave desert in California (Miller and Franklin, 2002) used generalized linear models as well as classification trees to predict the presence of four vegetation alliances in the Mojave desert in California using climatic and topographic variables (elevation, landform). The results showed that the predictions between the two techniques varied
significantly. Classification trees had in general higher accuracy than GLMs and that the predictions were better for those alliances for which there were more data available.

Modelling within a GIS has also been employed to investigate the potential effects of rapid anthropogenic climate change on both species and community distribution. Using climate response surfaces, Huntley et al. (1995) modelled the potential future ranges of eight higher plants in Europe. The study suggests that macroclimate parameters are correlated with the distribution of all eight species at European scale. Brzeziecki et al. (1995) examined the spatial distribution of forest communities in Switzerland under potential climate change. The approach used an empirical vegetation-site model to provide information about current communities’ distribution. A climate-sensitive model was subsequently employed to predict future distributions for two different climate scenarios. The comparison of the results showed that an increase in temperature of 2.2 - 2.75°C is likely to cause up to 70% changes in the forest distribution if compared to an increase of 1.1-1.4°C.

Another tool frequently employed in vegetation studies, which has an important role to play, particularly in ecological sensitive areas such as the Amazon (e.g. Alves et al., 2003) is remote sensing. Based on the spectral reflectance of different vegetation types, the use of multispectral images provide a means for vegetation classification and mapping (Salvador and Pons, 1998; Lillesand and Kiefer, 2000). The basic features and suitability of satellite imagery as a source of data for vegetation mapping have been discussed in Scott et al. (2002) as well as Alexander and Millington (2000). In the past the main criticism to the use of satellite imagery and aerial photography was that neither of these had accomplished the level of detail often needed in vegetation studies (Kalliola and Syrjänen, 1991). Therefore most of the vegetation studies based on remotely sensed data have performed analysis of the structure and physiognomy of vegetation rather than its floristic composition. However, the present generation of high resolution satellite systems (e.g. IKONOS) generates imagery with nominal resolutions between 1-5 m. This will substantially improve the capability for vegetation and habitat mapping (e.g. Reed 2003). The spatial detail available will be supplemented with the advantages of satellite imagery such as repeat viewing, large area coverage, digital format and good geometric properties (Griffiths et al., 1999).
There is a wide range of ecological applications where remote sensing has been used directly for data derivation or indirectly for validation and comparison with datasets derived from existing maps alone (see review in Skidmore, 2001). Some common applications include looking at ecosystem dynamics Ranson et al. (2001), forest impacts on vegetation (Bucini and Lambin 2002) and habitat mapping (Oindo et al., 2003).

3.1 The role of Digital Elevation Models

The influence of topography on vegetation patterns is well documented (Franklin, 1995) therefore the use of topographic attributes such as elevation and slope derived from a Digital Elevation Model are among the most common variables employed in vegetation modelling studies (e.g. Tappeiner et al., 1998). In particular, DEM topography gives better results for community based approach rather than species based approach. As demonstrated by Zimmerman and Kienast (1999) for alpine grasslands in Switzerland, the main reason for that is that species-realised niches are too complex to be modelled using DEM derived topography.

A digital elevation model (DEM) is any digital representation of the continuous variation of relief across space (Burrough and McDonell, 1998). DEMs have many geomorphological, environmental and pedological applications (Moore et al., 1991; 1993). Today, increasing use is being made of digital elevation models (DEMs) within a GIS to obtain a range of terrain attributes. The terrain attributes derived from a DEM can be classified into primary and compound (Moore et al., 1991). The former refer to elevation, aspect and slope while the latter to potential solar radiation, soil properties as well as temperature.

The importance of topographic parameters such as slope, aspect and elevation in determining vegetation composition and distribution has been highlighted by many ecological and silvicultural studies (Franklin, 1995). Therefore, primary terrain attributes derived from DEMs have been used extensively in predictive vegetation mapping studies (Zimmerman and Kienast,
1998; Miller and Franklin, 2002), as well as predictive maps of wildlife habitats (Gurnell et al., 2002).

The main data sources for most DEMs are: ground surveys, maps, satellite images and aerial photographs. Global Positioning Systems (GPS) have also been used recently to provide supplementary height data in order to improve model representation of breakline features (Stocks and Heywood, 1994). The accuracy of the elevation data contained in a DEM depends on the method used (Table 2).

**Table 2** Data sources and DTM accuracy (modified from Stocks and Heywood, 1994)

<table>
<thead>
<tr>
<th>Data source</th>
<th>Capture method</th>
<th>DTM accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air survey</td>
<td>Stereo plotters</td>
<td>High</td>
</tr>
<tr>
<td>Satellite image</td>
<td>Stereo autocorrelation</td>
<td>Moderate</td>
</tr>
<tr>
<td>Maps</td>
<td>1. Digitising</td>
<td>Low-Moderate</td>
</tr>
<tr>
<td></td>
<td>2. Raster scanning</td>
<td></td>
</tr>
<tr>
<td>GPS</td>
<td>Direct digital output from GPS field station</td>
<td>High</td>
</tr>
<tr>
<td>Ground survey</td>
<td>Direct digital output from theodolite field station</td>
<td>Very High</td>
</tr>
<tr>
<td>Synthetic Aperture Radar</td>
<td>Interferometry</td>
<td>Very High</td>
</tr>
</tbody>
</table>
4. **Empirical Models for Vegetation Mapping**

The efficient assessment of biodiversity requires a wide variety of biological, ecological and cultural information. Even in countries like the United States and the United Kingdom, where this information is readily available, it is generally not acquired as part of a co-ordinated environmental management system. As often pointed out, this information is scattered among different organisations in incompatible formats making their location and integration problematic (Griffiths et al., 1999). The CORINE and NATURA programmes established by the European Community are an important step towards improved nature conservation. The first refers to the design, compilation and use of an inventory of sites of major importance for nature conservation in the European Community (Council of Europe 1985). The second is the recent proposed network of protected sites within the European Union (Council of Europe, 1992). Despite the efforts described above, there are still important areas within Europe that form part of this network, such as Lefka Ori (see case study below), but where information on species and community distribution is limited.

Efficient decision-making in nature conservation is based on the availability of timely and accurate information. Traditionally the method for deriving this information has been field survey. However, both time and limited resources (i.e. human resources and money) limit the scope of surveys. For example, survey data are inevitably time and location specific. Another limitation on survey results can be imposed by the remoteness of the study area.

An approach with real potential to solve these problems is empirical modelling. The generic steps of empirical modelling is shown in Figure 2. The first step is a field survey of the entities in question and their environment. Then statistics is usually employed to establish a relationship between the entity (community, species) and some measurable (mappable) environmental variables. This results into a statistical model which is given spatial dimension within a GIS and applied to a wider area.
Empirical approaches to prediction are based on the assumption that previously observed relationships will continue to hold in the future (Elston and Buckland, 1993), i.e. they are static models. There are many successful applications of this approach to both habitat and vegetation modelling and mapping. For instance, Buckland and Elston (1993) demonstrated a case of empirical modelling in north-east Scotland. The models derived from literature, as well as site data for green woodpecker, redstart and red deer, were integrated with a GIS to measure habitat suitability, thus predicting the species’ distribution. Another study on the Mojave desert in California employed Generalised Linear Models and classification trees to predict the presence of four alliances (Miller and Franklin, 2002) based on topographic and climatic variables. Moreover, an example from the Mediterranean is the study by Carmel et al. (2002) where logistic and linear regression were used to develop an empirical model of vegetation dynamics for an area in the Galilee mountains, Northern Israel using topography, disturbance maps (such as grazing logging and fire) and vegetation properties.

Since empirical models are static, they perform better when used in a species-based approach. The main reason is that species assemblages are more transient than species in geological time, as
palaeoecological evidence indicates (Franklin, 1995). There are also a number of practical considerations that have to be taken into account when employing an empirical approach. First, how applicable is the modelling procedure in a different area or data set, i.e. extrapolation? In other words, how representative is the model for unsampled areas? It has been reported that when a model was applied to an area other for which it was developed its performance was reduced (e.g. Carmel et al., 2002). Empirical models relating vegetation composition to measured site variables are based on ground samples. These samples of predetermined area, located subjectively in order to represent ideal vegetation types, inevitably comprise a small part of the mapped region (Davis and Goetz, 1990).

Empirical modelling is not explanatory but rather a series of statements about the relationships between the presence of a species or community type with the input parameters. Thus, empirical models are dependent on the selection and measurement of environmental factors. Problems result from inter-relationships between variables, while accuracy tends to decline with the number of input variables and the complexity of the environment. These factors have to be measurable and mappable if the results are to be integrated within a GIS. When GIS variables are derived from other variables, errors may propagate from one stage to the next. Finally, important variables may be excluded when there is insufficient information, leading to the use of surrogate variables, which may not represent the original variable accurately (Davis and Goetz, 1990). For example, studies on vegetation modelling have concluded that the use of altitude is probably insufficient to reveal the clear influences of climatic variables on vegetation (Tappeiner et al., 1998). Similar studies on the Central Pyrenees (Del Barrio et al., 1997) and the Central Alps (Tappeiner et al., 1998) suggested that the distribution of dwarf shrub communities is influenced by snow cover which controls the length of the growing season and the uppermost limit of continuous plant cover in alpine environments.

Another important consideration related to variable selection when extrapolating at a landscape level is the effect of the spatial scale of the study. Generally the number of important variables decreases as the scale of the study becomes coarser. This is due to the fact that some variables change more than others when the scale changes (Mentemeyer and Box, 1987). At coarse scales the abiotic factors take over from the biotic factors that seem to prevail at finer scales (Olsvig-
Whitakker, *et al.*, 1992). When extrapolating species distribution in space, direct gradients (i.e. those that regulate physiological processes but not consumed by plants) or their surrogates should be used rather than indirect gradients (i.e. those that have no direct influence on plant growth but are correlated with resources and regulators). The reason being that the latter, according to Austin and Smith (1989), are complex and location specific. Although physical factors may explain species richness patterns better than history, as demonstrated by Birks (1996) for the flora of the Norwegian mountains, it has been suggested that the inclusion of disturbance history when mapping actual or potential vegetation, is necessary (Franklin, 1995). Where a long series of records have been available disturbance factors has been included in the modelling process (e.g. Carmel *et al.*, 2001). However, it is generally admitted this type of information is rarely available for incorporation into a GIS (Norton and Nix, 1991).

4.1 **Case Study: Lefka Ori Crete**

Lefka Ori is the most imposing of three mountain massifs in Crete (Figure 3) with the highest local and regional endemism rate among the Greek mountains (Strid 1995) hosting 22 threatened species (Phitos *et al.*, 1996). The remoteness and rugged terrain of the massif though impose limitations to field surveying (Figure 4). Therefore up to now and despite its ecological importance it has not been mapped in detail.

The aim of the study was to demonstrate the potential of an empirical modelling/GIS approach to provide data for conservation and management. The adoption of this approach was also due to the time constraints and the limited resources (i.e. manpower) in this study. Since adequate information for many of endemic and rare species is not available, adopting the species modelling approach was not feasible. Hence, a community modelling approach was employed instead. A predictive vegetation model was established (Figure 5) and verified for plant community distribution in an area of the alti-mediterranean zone of the Lefka Ori massif, using GIS (Vogiatzakis, 2000).
Figure 3. Map of the case study area: Lefka Ori Crete.

Figure 4 West Pachnes (2347m) summit Lefka Ori, Crete, Greece. View from the Northeast
Figure 5  Research steps for predictive vegetation mapping in Lefka Ori, Crete
Based on earlier work on the distribution of plant communities on the massif (Vogiatzakis et al., 2003) three variables were identified as significant determinants of plant community distribution, namely altitude, slope and geomorphology. The range of values for each of these variables (two continuous one categorical) was split into discrete classes and these classes were tested (against four community types) with Binary Logistic Regression to provide constraints/decision rules for the construction of a spatial model. The constraints were then used to produce a model for every community type. The model was implemented within a Geographical information System (GIS) to predict the distribution of each community type in the selected study site (Figure 6). The evaluation of the model using an error matrix gave an overall accuracy of 79% demonstrating that the variables incorporated in the model were important. However, there have been locations where no prediction was made and in transitional areas the improvement of the mapping quality of surface cover types is of considerable importance for improving the predictive accuracy of the models.

**Figure 6** Predicted distribution of *Dianthus - Lomelosia* community type in the Central Lefka Ori, Crete Greece
5. Conclusion

The complexity of nature and in particular the variation of species interactions and their responses to the environment make predictive models a difficult task (Hobbs and Morton 1999). The conceptual issues, methods and limitations of these models have been thoroughly reviewed by Franklin, (1995), Guisan et al. (2000) and Scott et al. (2002). During the last 30 years parallel developments in computer based cartography and modelling have improved our abilities to predict species and/or community distribution and therefore extend ecological research whether for nature conservation or planning applications. The main and indispensable tool for this purpose is a GIS due to its ability to merge data from different sources (i.e. field surveys, printed or digitised maps, remotely sensed imagery), construct spatial models or incorporate models build externally.

Moreover, other tools of spatial analysis are gradually being built into GIS environments providing new insights in ecological problems. For example, geostatistical methods for optimal interpolation have great potential for ecological applications. Although primarily used in soil science (Goovaerts, 1999; Webster and Oliver, 2001) geostatistics have been used to determine the spatial relationships between canopy openness and seedling performance in a secondary lowland forest in Borneo (Bebber et al. 2003), to examine the effects of migratory grazers on spatial heterogeneity of soil nitrogen properties in a grassland ecosystem (Augustine and Frank 2001) and to determine the patterns of diatom distributions at Lake Lama, Central Siberia (Kienel and Kumke, 2002). The use of geostatistics in combination with GIS has been advocated by Burrough (2001) and has been facilitated by the current availability of geostatistical routines/extensions within GIS software such as Idrisi and ArcGIS.

In addition to geostatistics most GIS packages have also inbuilt routines for fuzzy set thery (Zadeh, 1965). Fuzzy sets are inexactlty defined classes that characterise an attribute or phenomenon that for various reasons does not have sharply defined boundaries (Burrough and McDonnell, 1998). Fuzzy methodology has become an invaluable tool when dealing with spatial uncertainty in ecology (see review by Hunsaker et al., 2001). Moreover, it is a critical issue for spatial analysis and modelling since both GIS and remote sensing create approximate representations of geographical objects (Goodchild, 1994; Foody 2002). Examples in ecological
research include the use of fuzzy logic for classification of ecological data (e.g. Boyce 1998, Eyre et al. 2003), land use and land cover (e.g. Burrough et al., 2001) and for mapping transitional areas (Kent et al., 1997; Armitage et al., 2000; Townsend and Walsh, 2001).

All of these tools have facilitated modelling approaches such as empirical modelling which despite its limitations as described herein, can overcome the problems imposed by traditional field surveys.
References


