Using GIS and remote sensing for ecological mapping and monitoring

Jennifer Miller* and John Rogan†

*Department of Geography and the Environment, University of Texas at Austin, TX, USA
†Graduate School of Geography, Clark University, Worcester, MA, USA

10.1 Introduction

The ability to map and monitor ecological phenomena over large spatial extents has become a focus of renewed research in the context of increasing awareness of human activities and environmental change (Busby, 2002; McDermid et al., 2005; Liu and Taylor, 2002). Human activities substantially impact most of the terrestrial biosphere, currently at rates and spatial extents far greater than in any other period in human history (Kerr and Ostrovsky, 2003). Numerous organizations, disciplines and initiatives have formed in the last 15 years in response to the myriad challenges to sustainable resource management and ecological protection, e.g. the International Association of Landscape Ecology, the NASA Land Cover/Land Use Change Program. These interdisciplinary and integrative initiatives agree that scientifically sound and sustainable resource management requires ecological data of variable spatial and temporal characteristics to provide the scientific understanding required to measure, model, maintain and/or restore landscapes at multiple scales (EPA, 1998; Wiens et al., 2002). Research efforts in support of sustainable ecosystem management have focused on characterizing ecosystem condition and change, exploring the effects of different management schemes, and understanding how natural and anthropogenic processes affect ecosystem functioning (EPA, 1998).
Solutions to these problems require spatially explicit, timely, ecological data, often combined with statistical models in a geographic information system (GIS).

Current research illustrates how ecological problems ranging from biodiversity loss to land-use change have benefited greatly from advances in geospatial technologies such as GIS and remote sensing, both in the provision of data and access to spatial data analysis tools. The integration of GIS and remote sensing for ecological mapping and monitoring, while addressed in earlier research (Stoms and Estes, 1993; Franklin, 1995; Goodchild, 1994), has become even more important as these data and technologies continue to evolve, and as ecological issues become more critical. The key motivations for integrating GIS and remote sensing for ecological research and management are:

1. The acceptance of the landscape context and scale for sustainable ecosystem management (Liu and Taylor, 2002).
2. The importance of retrospective and prospective monitoring for conservation (Urban, 2002; Turner et al., 2003).
3. Increased familiarity with GIS and remote sensing data and methods within resource management agencies (Jennings, 2000).
4. Improved geospatial data quality and availability (at reduced cost) (Rogan and Chen, 2004).
5. Reported advantages of using different types of geospatial data (from both GIS and remote sensing) for mapping and monitoring applications (Rogan and Miller, 2006; Zimmermann et al., in press).

Although the benefits of integrating GIS and remote sensing data for more effective ecological mapping and monitoring are many, the time, money and expertise required to take full advantage of the technology can be initially daunting. The information used for scientifically valid ecological mapping and monitoring needs to be frequently updated, sufficiently detailed and spatially continuous. Ecological inventories have historically been conducted through field survey – a time-consuming and expensive endeavour, particularly when study sites are large and/or remote, and when long-term monitoring is a concern to resource managers (Rogan and Chen, 2004). This field work paradigm has implicitly affected both the typical study area size and the spatial scale of observations associated with ecological research. Field data are also typically collected based on some purposive sampling scheme, in which information on a specific ecological attribute (e.g. species abundance, timber inventory) is of primary interest, and therefore may not be appropriate for describing other attributes of subsequent interest (e.g. productivity, fuel loadings, habitat suitability). Lack of familiarity and background knowledge, equipment cost and complexity of data-processing methods are often cited as factors that prevent even wider use of remote sensing approaches by ecologists as well as by practitioners in other disciplines (McDermid et al., 2005; Treitz and Rogan, 2004).
10.1 INTRODUCTION

The availability of ecological datasets, collected through remote sensing (e.g., land-cover, NDVI) or derived within a GIS (e.g., topographic moisture index, incoming solar radiation) at local to global scales, has revolutionized the way ecological research is conducted (Cohen and Justice, 1999; Rushton et al., 2004). GIS enhances the ability to derive information from remotely sensed data, and remotely sensed data can describe actual environmental conditions for expedient updating of GIS databases. The synoptic perspective, temporal frequency and repeatability of remotely sensed measurements have been invaluable for detecting and monitoring change (Rogan et al., 2003).

In a recent review of remote sensing applications in ecological research, Kerr and Ostrovsky (2003) identified three main application focus areas: land-cover classification, integrated ecosystem measurements and multitemporal change detection. This chapter examines the ways in which remotely sensed data have been integrated with GIS data and modelling approaches in the context of these three areas. We focus on species distribution models (SDM) and biodiversity mapping/modelling as particular cases of ecological mapping, we summarize the GIS and remotely sensed environmental data that are most commonly used in these applications, and we include a case study that integrates GIS and remote sensing for ecological monitoring (land-cover change mapping). Although the traditional inconsistency in spatial scale between remotely sensed data (indirect) and ecological field observations (direct) has been a major obstacle to more extensive integration of remote sensing in ecological research, access to increasingly fine spatial resolution data has resulted in great progress in this area (Turrer et al., 2003; Aplin, 2005; Kerr and Ostrovsky, 2003). Further, advances in remote sensing theory, data and technology over the past 35 years have led to general and robust methods of large area data collection that, for many ecological attributes, can provide more reliable estimates than field methods (Davis and Roberts, 2000). The integration of GIS and remotely sensed data and techniques can greatly facilitate all steps of data collection, compilation, analysis and visualization. The potentially synergistic benefits of integrating GIS and remotely sensed data with statistical methods are still being explored and identified.

The three main sources of information used in ecological mapping and monitoring applications are shown in Table 10.1. Field observation provides the most detailed and fine-scale information, although the spatial coverage is not continuous. Field data are also expensive and time-consuming to collect, and many of the observations are relatively subjective or suited for a narrow purpose. GIS data can provide continuous spatial coverage (usually through interpolation methods), albeit at coarser resolution. Although we use the term ‘species’ here, these models can also be used to predict species assemblages (Ferrier et al., 2002; Franklin, 1995).
Table 10.1 Characteristics of field, GIS and RS data sources used for ecological mapping and monitoring

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Benefits</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field observation</td>
<td>Fine spatial scale, Detailed information, Direct observation</td>
<td>Limited temporal extent, Incomplete spatial coverage, Expensive, Subjective</td>
</tr>
<tr>
<td>GIS</td>
<td>Associated with potential distributions, Can be used for species-level mapping, Can be used to derive direct and resource gradients</td>
<td>Limited spatial resolution, Unknown accuracy, Cannot be frequently updated, Pixel values usually result from interpolation, Indirect gradients most readily available</td>
</tr>
<tr>
<td>RS</td>
<td>Associated with actual distributions, Allows data collection in remote areas, Synoptic perspective, Systematic measurement for every pixel; Complete spatial coverage, Enable larger study areas, Multitemporal; high temporal resolution, Cost-effective for large extents</td>
<td>Atmospheric obstructions possible, Expensive for fine spatial scales, Less detailed information, Processing methods intimidating (to untrained users), Usually represent indirect or functional gradients</td>
</tr>
<tr>
<td>GIS and RS (integrated)</td>
<td>Data can be upscaled, More consistent and objective databases, Can provide updated environmental data, Data are readily available, Direct, resource, indirect, and functional gradients can be combined</td>
<td>Compounding of quantitative–positional errors, Lack of automated methods to aid integration, Paucity of raster GIS data with fine spatial resolution (to match RS data)</td>
</tr>
</tbody>
</table>

Spatial resolution and lower or unknown positional accuracy. GIS data, particularly digital elevation models (DEM), have been used to derive complex environmental variables that are more ecologically relevant (e.g. topographic moisture index, potential solar radiation). Remote sensing facilitates data collection in difficult- or impossible-to-reach areas and provides an important synoptic and multitemporal perspective. Remote sensing also provides a systematic value for each pixel and
10.3 GIS DATA USED IN ECOLOGICAL APPLICATIONS

spatially continuous coverage. However, fine spatial resolution imagery consistent with the scale of field observation is expensive, some amount of processing is required, and atmospheric obstructions can be problematic. The integration of GIS and remotely sensed data improves upon many of their individual limitations.

10.2 Integration of GIS and remote sensing in ecological research

Rogan and Miller (2006) summarized four ways in which GIS and remote sensing data can be integrated: (a) GIS can be used to manage multiple data types; (b) GIS analysis and processing methods can be used for manipulation and analysis of remotely sensed data (e.g. neighbourhood or reclassification operations); (c) remotely sensed data can be manipulated to derive GIS data; and (d) GIS data can be used to guide image analysis to extract more complete and accurate information from spectral data.

Remotely sensed data and techniques have been widely available since the early 1970s but the most common ecological application for which they are combined with GIS is still mapping land cover (Stoms and Estes, 1993). The coarsened scale of these land-cover maps, as well as the potential for circular reasoning when using them to model plant species distributions, renders them unsuitable for most SDM studies (Zimmermann et al., in press). The continuous properties of image spectral values and vegetation indices have seldom been used as predictor variables in SDM, although both show great potential (Frescino et al., 2001; Osborne et al., 2001; Suárez-Seoane et al., 2002; Zimmermann et al., in press). GIS has been a mainstay of SDM through the derivation of and analysis with bioclimatic factors associated with species distributions (see Franklin, 1995; Guisan and Zimmermann, 2000) but other factors, such as competition and disturbance, may be more appropriately described by remotely sensed data. Despite the synergistic potential of combining GIS and spectral data, remote sensing is rarely used directly in ecological mapping studies (but see section 10.5).

10.3 GIS data used in ecological applications

We use ‘GIS data’ here to describe non-spectral digital environmental data, as they are stored, manipulated and typically derived in a GIS. These data are derived either by interpolating field or station observations to a continuous surface (e.g. temperature) or by calculating new surfaces from existing spatially continuous data (e.g. slope from a DEM). The availability of digital data that represent increasingly complex environmental characteristics provides the basis for SDM (for reviews, see Franklin, 1995; Guisan and Zimmermann, 2000; Guisan and Thuiller, 2005). Climatic and topographical variables are the most widely used predictors in SDM,
as they describe broad-scale physiological tolerances related to water and temperature, and site energy and moisture availability associated with micro-climates, respectively (Franklin, 1995). These environmental variables are reviewed below, along with a summary of gradient analysis, which provides the conceptual framework for describing the way in which environmental gradients influence species distributions.

10.3.1 Gradient analysis

SDM evolved from research methods that used gradient analysis to explore how plant species composition and distribution change along environmental gradients (Whittaker, 1973; Kessell, 1979; Franklin, 1995). The range of environmental conditions a species is physiologically able to tolerate defines its fundamental niche, which is analogous to its potential distribution. Due to other factors, such as competition or disturbance, a species typically only occupies a subset of its fundamental niche, which is termed its ‘realized niche’, or actual distribution. SDM involves the quantification of the species–environmental gradient relationship, the result of which is a species habitat distribution map. The nature of the gradient variable determines how robust the resulting model is likely to be. Austin (1980) describes three types of environmental gradients:

1. **Direct gradients** are those in which the environmental variable has a direct physiological effect on species growth but is not consumed (e.g. temperature).
2. **Resource gradients** have a direct physiological effect on growth, and are actually used or consumed (e.g. water, nutrients).
3. **Indirect gradients** have no direct physiological effect and are likely the result of a location-specific correlation with one or more direct gradients (e.g. correlation between temperature and elevation makes elevation an indirect gradient).

Data describing indirect gradients are usually more readily available in digital format, or easily measured in the field, and are often highly correlated with observed species patterns (Guisan and Zimmermann, 2000). However, these relationships are location- and gradient-specific, usually describing a combination of other direct or resource gradients, and models in which they are used are generally not appropriate for extrapolation beyond the area in which the data were collected (Franklin, 1995; Austin and Gaywood, 1994). Variables that represent direct or resource gradients are the most suitable for extrapolating across space and time, although Franklin (1995)

---

2 Recent articles have discussed the ambiguity in the use of the word ‘habitat’ (de Leeuw et al., 2002; McDermid et al., 2005). Here we consider habitat to refer to the type of environment (as measured by a suite of environmental factors) in which an organism normally occurs.
10.3 GIS DATA USED IN ECOLOGICAL APPLICATIONS

points out the incongruity of using empirical data on actual species distribution (realized niche) to map potential species distribution (fundamental niche).

A fourth type of gradient added by Müller (1998), ‘functional’ gradients, is associated with species response to direct and resource gradients. Many of the spectral-derived variables used in ecological applications of remote sensing, such as productivity, biomass, and leaf area index, can be described as functional gradients. Models that include functional gradients allow for accurate depiction of actual landscape composition, structure and function (Rollins et al., 2004) and may be particularly useful for modelling disturbed landscapes (Lees and Ritman, 1991; Frescino et al., 2001). Functional gradients should be considered location-specific in the same sense as indirect gradients, although Zimmermann et al. (in press) discuss an exception involving the use of spectral-derived variables that describe vegetation structure as direct gradients for modelling bird species distribution.

The distinction between mapping actual vs. potential vegetation species has been discussed previously (Woodcock et al., 2002; Franklin, 1995). Maps of potential species distribution are used in models to investigate effects of climate change on vegetation distribution, and as the first step in parameterizing dynamic vegetation models (Guisan and Zimmermann, 2000). However, actual species distribution maps may be more appropriate for certain applications, such as resource management or biodiversity measurement. Figure 10.1 illustrates the typical flow of information and data types used when GIS and remotely sensed data are integrated for ecological mapping and monitoring. A SDM based solely on GIS data, such as bioclimatic variables, makes the often untenable assumption that species distributions are at equilibrium with their environment (Guisan and Zimmermann, 2000). The resulting potential habitat distribution maps, particularly for species whose distributions have been modified by anthropogenic effects, are likely to be over-predicted, as they describe the fundamental rather than the realized niche (Thuiller et al., 2004). Coarse-scale predictor variables, such as climate, are typically correlated with potential distributions, while finer-scale topographic variables are associated with actual distributions (Thuiller et al., 2003; Franklin, 1995). Guisan and Thuiller (2005) further divide the hierarchical effects into (from global to local scale) limiting climatic factors, dispersal factors, disturbance factors and resource factors. Additional information on biotic and abiotic factors that reduce a species’ fundamental niche must be included in the SDM to produce a map of actual species distribution. Functional gradients that describe actual ecosystem characteristics, such as Normalized difference vegetation index (NDVI), can be used as predictor variables to produce actual species distribution.

While many direct and resource gradients important to species distributions are still unknown, immeasurable or difficult to characterize across a landscape (Whittaker, 1973), the ability to derive increasingly complex (and more ecologically relevant) environmental variables using GIS and remote sensing has great potential. Accurate and timely spatial information describing actual ecological characteristics is essential for predicting future conditions, and remotely sensed data
are particularly useful for this purpose. The next generation of gradient models will incorporate direct, resource and functional gradients to more accurately map ecological characteristics (Franklin, 1995; Austin and Heyligers, 1989).

10.3.2 Climate

Climate has been linked with vegetation distribution from at least the early 1800s, when von Humboldt wrote about the relationship between latitude and vegetation type (as cited in Jongman et al., 1995). Early global plant distribution maps based on climatic factors alone were surprisingly accurate (e.g. Holdridge, 1947, 1967) and the relationship between climate and vegetation remains very important in broad-scale vegetation modelling. In addition to providing the fundamental relationship on which static vegetation maps are based, an understanding of the complex feedback relationship between vegetation and climate variability is necessary to parameterize dynamic vegetation models used to study global biogeochemical cycles.

Climate data are particularly important in SDMs used for predicting consequences of global warming on plant and animal distributions. Many plants and animals are limited in their distribution by temperature extremes. Teixeira and Arntzen (2002)
observed that slight changes in climate will have a particularly important effect
on ectothermal (cold-blooded) animals, as temperature directly affects most of
their physiological processes, as well as other important environmental conditions
(air humidity, soil moisture, and vegetation composition). Thuiller et al. (2003)
found that temperature and precipitation extremes were effective surrogates for
bioclimatic factors with more direct physiological roles (e.g. evapotranspiration)
in limiting the distribution of Mediterranean vegetation in Spain. While climate
variables, such as temperature and precipitation, tend to have broad-scale influence
on species distributions, many finer-scale bioclimatic indices, such as potential solar
radiation, mean relative humidity and potential evapotranspiration, can be derived
using other GIS data (e.g. elevation, aspect) and may be more directly related to
species distributions (Franklin, 1998; Cairns, 2001; Leathwick, 1998; Meentemeyer
et al., 2001).

Thuiller et al. (2004) found that climate was the most important driver of species
distributions (plants, mammals, birds, reptiles and amphibians) in Europe. They
observed that, at the relatively coarse scale of their study (50 km), land-cover
provided largely redundant information relative to climate. However, using finer-
scale land-cover data (1 km) to predict plant species presence in Britain, Pearson
et al. (2004) found that availability of suitable land-cover was more important than
availability of suitable climate. Cumming (2002) suggested that the seasonality
of precipitation may be as important as its magnitude, particularly for animal
distributions. Climate can influence animal distribution both directly (e.g. bird
migration in winter) and indirectly (e.g. land cover and food availability) (Venier
et al., 2004). Although their importance in determining species distributions is well
established, digital climate surfaces have generally been produced by interpolating
ground station data and are of limited availability and quality for many areas (Parra
et al., 2004). In a recent study, Suárez-Seoane et al. (2004) explored the use of
climate variables derived from METEOSAT in a model of bird distribution in Spain.
They found that these 5 km resolution data showed great potential as an alternative
to interpolated climate surfaces, particularly in areas where meteorological stations
were sparse.

10.3.3 Topography

While bioclimatic factors, such as water availability, temperature and insolation,
are the main drivers of species distributions, topographic variation modifies their
influences, resulting in increased spatial heterogeneity associated with microclimatic
effects. Although simple topographic variables that represent indirect gradients,
such as elevation, slope and aspect, should have less influence when used along
with direct and resource gradients in models, they are often empirically important,
as they tend to be derived with higher accuracy (Rollins et al., 2004; Guisan and
Zimmermann, 2000). One general assumption has been that as the processing steps
involved in deriving a topographic variable increase, so too does its susceptibility to error (Guisan and Zimmermann, 2000), although Van Niel et al. (2004) note that this is not always the case. In a study that simulated error propagation in the derivation of topographic variables, they found that in some cases more complex variables, such as net solar radiation, were less affected by error than comparatively simple variables, such as slope and aspect (Van Niel et al., 2004). In a similar study, Holmes et al. (2000) found that topographic variables derived by compounding values from a large number of other grid cells were more error-prone, and that while global error estimates may be low, local error could be quite high.

Derived topographical variables used in SDM include potential solar radiation, landscape position, slope curvature and topographic moisture index (for reviews, see Florinsky, 1998; Franklin, 1995; Moore et al., 1991). Potential solar radiation can be used to simulate a direct gradient that describes potential evapotranspiration and soil moisture. Landscape position, the location of a grid cell relative to surrounding grid cells ( upslope or downslope) is related to a combination of soil properties, specifically depth, texture and potential moisture ( Franklin, 1995). Slope curvature and topographic moisture are both related to the water availability of a site (Moore et al., 1991). In a study relating mammalian species richness to environmental variables, Tognelli and Kelt (2004) used elevation range for each quadrant as a proxy variable to represent habitat heterogeneity. White et al. (2005) found that topographic variation influenced the way in which vegetation responded to interannual climatic fluctuations (e.g. ENSO), although it is often not explicitly addressed in climate models. Topography also influences the onset, rate, pattern and duration of disturbance intensity and severity (Rogan and Miller, 2006).

Other GIS variables, such as geology and soil type, can also be used to represent moisture and nutrient availability, although usually at a coarser scale. GIS variables, such as ‘distance to _____ (roads, water, edge, etc.)’, can represent proximity to disturbance or important resources (Osborne et al., 2001). Landscape metrics, such as pattern, structure and heterogeneity, can also be quantified and used as predictor variables (McGarigal and McComb, 1995; Gottschalk et al., 2005). Habitat suitability maps for other species can also be used to represent potential competition or predation, or can be used to stratify sampling schemes for more rare species (Edwards et al., 2005).

10.4 Remotely sensed data for ecological applications

Over the past 35 years, remotely sensed data have steadily become an invaluable information source for ecological characterization and survey (see recent reviews by Gong and Xu, 2003; Coppin et al., 2004; Aplin, 2005; Turner et al., 2003; Kerr and Ostrovsky, 2003; Rogan and Chen, 2004). This is primarily due to the effectiveness of air-borne and space-borne remote sensing platforms and sensors that facilitate observation of biophysical attributes over extensive areas at multiple
10.4 REMOTELY SENSED DATA FOR ECOLOGICAL APPLICATIONS

spatial, spectral and temporal scales (Stow, 1995; Jensen, 2000). Landsat, long considered the ‘workhorse’ of terrestrial remote sensing, has provided the longest-running time series of remotely sensed data at scales appropriate for regional studies. Thus, there is a perceived space–time mismatch between the information that ecologists want and what remotely sensed data can provide. Nonetheless, field data (ground reference) are often limited for regional-global scale investigations because they are rarely as widespread or timely as remotely sensed data (Pettorelli et al., 2005).

The last 8 years have witnessed a proliferation of satellite platforms with a large number of sensors (e.g. Terra and ENVISAT) and improved spatial resolutions (e.g. IKONOS-2 and Quickbird data have pixels that cover an area of 16 m² or less) that can also serve the needs of timely and cost-effective resource management (Franklin, 2001). Our focus is on ‘passive’ remote sensing systems, although we acknowledge the benefits of ‘active’ systems (see Davis and Roberts, 2000; Kasischke et al., 1997).

Remotely sensed data are used primarily in ecological research to characterize land cover, describe habitat structure and derive measurements of biophysical properties. The ability of remotely sensed variables to act as surrogates for important ecological characteristics (e.g. biodiversity, productivity) is a function of the closeness of the relationship between the measured radiation and the environmental variable of interest. ‘State’ variables are those that can be described directly by the measure of electromagnetic radiation, such as leaf area index and biomass (Curran et al., 1998; Curran, 2001). However, it is most often environmental variables that are indirectly related to the actual radiation measure, such as biodiversity and productivity, that are of interest to ecologists. There are a wide variety of options to choose from to exploit known relationships between values of optical/microwave data and the biophysical properties of ecological entities (Franklin, 2001; Davis and Roberts, 2000). The following section presents an overview of commonly used remote sensing data enhancements and key ecological targets typically investigated.

10.4.1 Spectral enhancements

Considering passive optical data only, numerous spectral transformation methods have been developed to concentrate and accentuate the biophysical signal from the surface into an enhanced spectral ‘feature’ (Roberts et al., 1998). Spectral vegetation indices (VIs) have been used since the late 1960s, with continued evolution of new types of VI and uses. Advances in technology for both spectral sensing platforms and analytical techniques have led to a wide range of applications for VI, ranging from evapotranspiration estimates to forest structure quantification (Wulder, 1998).
The Normalized difference vegetation index (NDVI) has become the most extensively used VI in ecological remote sensing. NDVI is a dimensionless spectro-radiometric measurement derived from optical remotely sensed data that is correlated to micro- and macro-level characteristics of plants (for review, see Pettorelli et al., 2005). It was initially developed as a measure of green leaf biomass (Tucker, 1979) but has also been used effectively in mapping other vegetation attributes, such as percentage cover, stem density, stand health, etc.

Since its introduction almost 30 years ago, the Kauth Thomas (KT or Tasselled Cap) image transformation has proved to be versatile in ecological remote sensing applications (see Kauth and Thomas, 1976; Crist and Ciccone, 1984). This transformation involves the statistical rotation of multispectral data space into a set of physically meaningful VIs that describe scene brightness, greenness and wetness. Brightness has positive loadings in all reflectance bands and corresponds to overall scene brightness, or albedo (Crist and Ciccone, 1984). Greenness, like many other correlates of vegetation amount (e.g. NDVI), is a contrast between the visible bands (especially Landsat TM band 3) and the near-infrared (Landsat TM band 4). Wetness presents a contrast of the visible and near-infrared bands (weak positive loadings) with the mid-infrared bands (strong negative loadings). Cohen et al. (1995) found that wetness was least sensitive to topographic variation, and therefore more powerful for predicting forest structural attributes. KT variables have been used in a variety of ecological applications, such as forest canopy mapping and selective harvest detection (Cohen and Fiorella, 1998).

10.4.2 Land cover

Spectral data have been used most often to derive some variation of a map of land-cover type (vegetation, biotype) or quality (biomass, NPP), from which habitat distribution (Osborne et al., 2001; Suárez-Seoane et al., 2002; Venier et al., 2004), measures of abundance (Luoto et al., 2002a) or biodiversity (Luoto et al., 2002b; Dębski et al., 1999; Waser et al., 2004; Tognelli and Kelt, 2004) are classified (see section 10.5). Land-cover mapping determines the current composition and distribution of landscape attributes, and this is subsequently used as the basis for assessing future change. Regions (Homer et al., 1997, Huang et al., 2003), nations (Cihlar et al., 2003), continents (Stone et al., 1994) and the globe (Hansen et al., 2000; Belward et al., 1999) have been mapped at various spatial resolutions with a range of remotely sensed data inputs. Satellite imagery efficiently provides information about vast areas and is, therefore, a useful tool for land-cover mapping across large extents. Neither aerial photography nor field data can provide equal amounts of information as efficiently (Franklin et al., 2003).

Digital image classification is a common approach for predicting the categorical class membership (e.g. forest type) of an observation (pixel), based on spectral band response values. Multitemporal and multisensor spectral measurements, along
with GIS data, can be used to add explanatory variables to an analysis that may help discriminate categories of interest (e.g., forest vs. grassland). Variations in the structural attributes of the forest stand may have a greater effect on the reflectance characteristics than tree species composition. Therefore, other mapped environmental variables associated with, or controlling, forest vegetation distributions, such as those related to terrain (digital elevation models), geology, soils, climate or land use, can be combined with image data in the classification process in various ways to aid forest type discrimination.

## 10.4.3 Habitat structure

A number of studies beginning in the 1950s have empirically and theoretically explored the relationship between pixel-level reflected absorbed and transmitted radiation and habitat structure characteristics. Most studies that estimate vegetation structural and biophysical parameters from remotely sensed data have used empirical methods to relate spectral data and various image derivatives to vegetation characteristics. If these parameters are strongly correlated with remotely sensed data, they can be used to predict those biophysical attributes over large extents (Woodcock et al., 2001).

Structural variables allow ecologists to better discriminate forest habitat types for conservation planning purposes. Based on the framework of Diamond (1988), habitat quality represents the diversity of resources as defined by habitat diversity and structural complexity (i.e., in both vegetative structure and types) (Stoms and Estes, 1993). Using remotely sensed data, forests are often characterized in terms of inventory parameters, which provide detailed data on the location, arrangement, distribution and pattern of forest resources (Wulder, 1998). Generally, inventory parameters such as vegetation type, canopy cover and canopy height are linked to habitat structure which has been recommended for assessing wildlife resources and planning conservation efforts (Roberts et al., 2000).

Remote sensing approaches have involved the characterization of canopy or crown cover (also described as crown closure – the vertical projection of vegetation onto the ground when viewed from above) with a good deal of success (Davis and Roberts, 2000). Authors have attributed the success of these studies to the dominance of canopy cover in the radiation ecology of their site environments. A variety of approaches (using passive and active remote sensing instruments) have been developed to map canopy cover because this variable has been a primary motivator of environmental/biophysical remote sensing research (Franklin, 2001).

Canopy height is an important hybrid variable in forest biodiversity studies (Davis and Roberts, 2000). Several researchers have reported relationships between canopy height and spectral measurements (Franklin et al., 1986; Cohen and Spies, 1992; Danson and Curran, 1993; Jakubauskas and Price, 1997). Variability caused by topographic variation is often problematic in most landscapes, resulting in imprecise
representations of vegetation structural properties (Franklin et al., 2003). In all cases, topographic effects were minimized either by stratifying the study area and forest type to zero-slope or by including a slope variable in regression models. Stand density is defined as the number of individual trees per unit area. Stand density is limited in its applicability to forest ecosystem studies, where measures such as cover, height and volume are more commonly required for habitat characterization and prediction (Davis and Roberts, 2000). Density, however, may be of greater importance in semi-arid and arid regions (more heterogeneous cover), where vegetation density and cover are more highly variable across the landscape (Franklin and Turner, 1992).

Remotely sensed data have been employed in the measurement of many structural variables other than those discussed above, such as stand volume (Ardo, 1992, Oza et al., 1996), basal area (Franklin et al., 1986; Danson and Curran, 1993) and fire fuels estimation (Stow, 1995; Cosentino et al., 1981). Several researchers have adopted comprehensive approaches to measuring forest stand structural variables and have either attempted to determine the total stand structural factor contribution to stand spectral response, or have evaluated the total stand structural information contained within various spectral vegetation indices (Danson and Curran, 1993; Jakubauskas and Price, 1997; Steininger, 2000). This approach can help ecologists learn more about the contributions of component hybrid variables to overall spectral response, rather than only knowing the relationship between spectral response and a single hybrid variable.

### 10.4.4 Biophysical processes

An increasing amount of attention is being paid to the representation of variables that represent ecological processes using remotely sensed data. Remotely sensed measures of ecological process and productivity include leaf area index (LAI) and net primary productivity (NPP) (Wulder, 1998). Productivity is seldom measured directly (in the field) but is estimated from associated variables, such as temperature, precipitation, solar insolation, actual and potential evapotranspiration, biomass or leaf area index (Davis and Roberts, 2000).

Leaf area index (LAI) is the standard expression for the leaf area of a plant community, defined as the total leaf area per unit ground cover. LAI is an important biophysical attribute of plants because of its potential as an indirect measure of vegetation canopy energy, gas and water exchanges (Chen and Black, 1992). Maximum LAI has been correlated with mean annual temperature, length of the growing season, mean annual minimum air temperature and water availability (Gholz, 1982; Wulder, 1998). Field-measured LAI measures are strongly correlated to VIs, especially NDVI (Chen and Guibeault, 1996). Unfortunately, the relationship between LAI and NDVI is frequently non-linear, and can be erroneously lower due to canopy shading in mature forest stands. Stratification of NDVI images by vegetation or
land-cover class is therefore often used for robust estimation over regional scales (Wulder, 1998).

Net primary productivity (NPP) is defined as the net flux of carbon from the atmosphere into green plants per unit time. NPP refers to a rate process, i.e. the amount of vegetable matter produced per day, week or year. Estimates of NPP are based on ecological models which require detailed inputs, many of which are feasible only when acquired using remote sensing (Wulder et al., 2004). Numerous studies have shown that NDVI is related to ecosystem function, particularly NPP (Friedl et al., 1994; Ramsey et al., 1995). Running (1990) incorporated surface temperatures derived from advanced very high resolution radiometry (AVHRR) with annually integrated NDVI to provide better estimates of NPP.

10.5 Species distribution models

SDMs have long been a staple in resource conservation and management efforts, as well as research on the effects of climate change. The most commonly used medium is a map of plant species distribution, which can subsequently be used to derive maps that show suitable habitat characteristics for particular animals (Scott et al., 2002; Franklin, 1995; Guisan and Zimmermann, 2000; de Leeuw et al., 2002; Woodcock et al., 2002). These SDM rely on the digital availability of important environmental variables that influence plant (and subsequently animal) distributions. The product of SDMs, habitat suitability maps, can be used to show current distributions, identify possibly suitable (potential) habitat currently unoccupied, and predict the probable effects of changing environmental conditions. Guisan and Zimmermann (2000) reviewed the increasingly large variety of statistical methods used to quantify the species–environment relationship and discuss some of the conceptual considerations important in method selection. Although basic statistical methods (e.g. linear regression) are now available as part of many GIS software packages, many of the assumptions they make about data (e.g. independent observations, linear relationships between response and predictors) are violated with biogeographical data. More sophisticated statistical analysis is usually done using dedicated or user-written statistical software (for recent overview, see Guisan and Thuiller, 2005). Austin (2002) observes that the continued use of inappropriate statistical methods for SDMs stems from a long-standing disconnect between the ecological knowledge of statisticians and the statistical abilities of ecologists. Austin (2002) notes that linear relationships are still often used in models that describe species–environment relationships, despite both ecological theory and empirical evidence that refute this (Austin, 1987, 2002; Bio et al., 1998). The use of remotely sensed data and the functional gradients they describe (Müller, 1998) in SDMs increases the need for more exploratory and flexible statistical methods.

Table 10.2 summarizes selected recent studies that have integrated GIS and remotely sensed data to model species distribution for inventory, atlas or biodiver-
<table>
<thead>
<tr>
<th>Study</th>
<th>Response</th>
<th>GIS data</th>
<th>Spectral data</th>
<th>Model*</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austin <em>et al.</em> (1996)</td>
<td>Bird presence</td>
<td>Topography, number of buildings, road length</td>
<td>Land cover, habitat heterogeneity (Landsat-5 TM)</td>
<td>GLM, discriminant function analysis</td>
<td>Scotland</td>
</tr>
<tr>
<td>Cumming (2002)</td>
<td>Tick species</td>
<td>Climate, topography, vegetation type, political regions</td>
<td>NDVI (AVHRR)</td>
<td>GLM</td>
<td>Africa</td>
</tr>
<tr>
<td>Dimböck <em>et al.</em> (2003)</td>
<td>Plant communities</td>
<td>Topography</td>
<td>NDVI, reflectance, texture (CIR photographs)</td>
<td>CCA</td>
<td>Austrian Alps</td>
</tr>
<tr>
<td>Franklin <em>et al.</em> (2000)</td>
<td>Plant communities</td>
<td>Topography</td>
<td>NDVI, BGW, red (band 3), near-IR (band 4), mid-IR (band 5) (Landsat TM)</td>
<td>CT</td>
<td>Southern California, USA</td>
</tr>
<tr>
<td>Frescino <em>et al.</em> (2001)</td>
<td>Forest presence, basal area, cover (%), density</td>
<td>Climate, topography, geology</td>
<td>NDVI (AVHRR), Land cover, red (band 3), near-IR (band 4), mid-IR (band 5) (Landsat-5 TM)</td>
<td>GAM</td>
<td>Utah, USA</td>
</tr>
<tr>
<td>Jeganathan <em>et al.</em> (2004)</td>
<td>Bird presence</td>
<td>Relative abundance of bushes, trees</td>
<td>NDVI, reflectance (Landsat-7 ETM+)</td>
<td>GLM</td>
<td>India</td>
</tr>
<tr>
<td>Study</td>
<td>Response</td>
<td>Predictor Variables</td>
<td>Model</td>
<td>Location</td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td>----------</td>
<td>------------------------------------------------------------------------------------</td>
<td>--------</td>
<td>----------------------------</td>
<td></td>
</tr>
<tr>
<td>Luoto et al. (2002)</td>
<td>Plant species richness</td>
<td>Topography, habitat structure, soil moisture</td>
<td>GLM</td>
<td>South-western Finland</td>
<td></td>
</tr>
<tr>
<td>Osborne et al. (2001)</td>
<td>Bird presence</td>
<td>Topography, proximity to disturbance</td>
<td>GLM</td>
<td>Spain</td>
<td></td>
</tr>
<tr>
<td>Parra et al. (2004)</td>
<td>Bird presence</td>
<td>Climate, topography</td>
<td>EE</td>
<td>Ecuadorian Andes</td>
<td></td>
</tr>
<tr>
<td>Pearson et al. (2004)</td>
<td>Plant presence</td>
<td>Climate</td>
<td>ANN</td>
<td>UK</td>
<td></td>
</tr>
<tr>
<td>Suárez-Seoane et al. (2002)</td>
<td>Bird presence</td>
<td>Topography, road, town, river distance and density</td>
<td>GAM</td>
<td>Spain</td>
<td></td>
</tr>
<tr>
<td>Thuiller et al. 2004</td>
<td>Plant, mammal, amphibian, reptile, bird presence</td>
<td>Climate</td>
<td>GAM</td>
<td>Europe</td>
<td></td>
</tr>
<tr>
<td>Venier et al. (2004)</td>
<td>Bird presence</td>
<td>Climate</td>
<td>GLM</td>
<td>Great Lakes, USA</td>
<td></td>
</tr>
<tr>
<td>Zimmermann et al. (in press)</td>
<td>Tree presence</td>
<td>Topography, climate</td>
<td>GLM</td>
<td>Utah, USA</td>
<td></td>
</tr>
</tbody>
</table>

*ANN, artificial neural network; CCA, canonical correspondence analysis; EE, environmental envelope; GAM, generalized additive models; GLM, generalized linear models; CT, classification tree.
sity applications. Climate and topography are the most often used GIS variables, and land cover and NDVI are the most often used spectral variables. Climate and NDVI are quite similar with respect to the biotic and abiotic factors they represent, although NDVI generally has higher spatial resolution and lower temporal resolution.

Venier et al. (2004) compared MSS-derived land-cover maps to those derived from AVHRR as predictor variables (along with climate) in a model to predict distribution of bird species. They found that, despite differences in spatial resolution, both MSS and AVHRR land-cover maps produced similar results, and that they could not discriminate between the direct and indirect (e.g. through land cover) effects of climate (Venier et al., 2004). Frescino et al. (2001) found that both classified and raw TM data resulted in better predictions of forest structure compared to AVHRR-derived NDVI. They suggested that classified spectral data, such as land cover, can provide more information than raw spectral data because they are discriminated by ecological characteristics and neighbouring pixels (Frescino et al., 2001). In models used to predict vegetation distribution in Southern California, Franklin et al. (2000) found that topographical variables were useful to discriminate between physiognomically similar (and therefore spectrally similar) chaparral species. NDVI has also been used to characterize habitat type by calculating a maximum value composite (MVC) over a 12 month period (Osborne et al., 2001) or as a surrogate for energy (Cumming, 2002). Principal components analysis of the MVC data has also been used to decompose the time series into a sequence of spatial and temporal components (Suárez-Seoane et al., 2002).

Some studies have investigated the relationship between environmental factors and vegetation indices by using VI as a response variable. In a regression tree model with greenness vegetation index (GVI) as the response variable, Michaelsen et al. (1994) found that vegetation type was the most important predictor variable, followed by burning treatment (burned areas experienced greater greenness). The relationship between elevation and GVI was more complex and they observed that, during the growing season, elevation acted as a surrogate for soil properties but that after May the relationship became inverse, possibly suggestive of temperature/moisture differences (Michaelson et al., 1994). In a model for the western USA, Stoms and Hargrove (2000) found that precipitation, temperature and available soil water capacity were the most important predictors of NDVI in undisturbed areas. Their model predictions deviated from actual NDVI values in areas of urban or agriculture land use (Stoms and Hargrove, 2000).

Landscape pattern and heterogeneity are also important predictors in SDM. In a habitat model for an endangered bird species in India, Jeganathan et al. (2004) found that density of bushes and trees, as measured on the ground and remotely, was the most important predictor variable. They compared ground-based bush/tree density information to that derived from satellite imagery and found that models based on the imagery performed better, although they suggest that ground-based surveys may be more directly relevant for habitat management decisions (Jeganathan et al., 2004).
Additional ‘spatial’ habitat variables can be subsequently derived from a digital map of habitat complexity/characteristics (Coops and Catling, 2002).

Statistical models that combined spatial and spectral data have also been used to produce maps for subsequent analysis, such as maps of fuel loads and fire regimes (Rollins et al., 2004) and to provide input maps for dynamic landscape simulation models (Franklin, 2002). Habitat suitability maps for several species have been combined to indicate possible levels of biological diversity within an area.

### 10.5.1 Biodiversity mapping

Biological diversity, or biodiversity, refers to variability within species, among species and in ecosystems. The applications discussed here are concerned with variability among species, also referred to as species richness. The ability to measure and monitor biodiversity, particularly critical in the context of environmental change (Nagendra, 2001), requires adequate, updated species inventories, as well as detailed knowledge of species–habitat relationships (Scott et al., 1993). The separate effects of climate change and land cover change on habitat loss worldwide are substantial. However, their interaction could be even more devastating (Travis, 2003).

SDMs have been used to produce habitat suitability maps for several different species, which have been combined and used as a surrogate of biodiversity (Ferrier, 2002). Availability of species data is often the limiting factor for this technique (Ferrier, 2002). Remotely sensed variables have also been used as a surrogate for biodiversity, most often by relating it to NPP, for which NDVI is used as a surrogate (Kerr and Ostrovsky, 2003; Aplin, 2005; Skidmore et al., 2003; Turner et al., 2003).

Skidmore et al. (2003) explain the correlation between NDVI and biodiversity as being a function of the debated diversity–productivity hypothesis in ecology. Increased availability of resources in highly productive ecosystems generally results in a greater number of species in a given area. Empirical evidence suggests that the relationship is actually unimodal, with maximum biodiversity associated with some optimal value of productivity, beyond which biodiversity begins to decrease. Skidmore et al. (2003) summarize possible explanations for this.

While others have used AVHRR NDVI data to confirm a unimodal relationship between species richness of plants and productivity (Oindo and Skidmore, 2002), Seto et al. (2004) did not find a quadratic model to be a significantly better fit than a linear model for bird and butterfly species richness in the Western USA. While they observed a positive correlation between bird and butterfly species richness and NDVI mean, maximum and standard deviation, they were unable to identify any functional relationships. However, they note that the relationships vary, based on scale, location and taxonomic group (Seto et al., 2004). Linear relationships may be observed empirically if the range of the species studied extends beyond the study area sampled, as Austin (2002) noted for species response to environmental gradients.
Seto et al. (2004) suggest that using NDVI as a proxy for species richness improves upon more subjective and time-consuming classification approaches, which require some \textit{a priori} (and often subjective) correlation between species richness and land cover. As a result, resource managers and ecologists may evaluate more quickly and efficiently the potential biodiversity impacts of alternative management and conservation strategies. Luoto \textit{et al.} (2002b) determined that environmental variables derived from a combination of satellite imagery (Landsat TM) and elevation data provided a reasonable surrogate for plant species diversity. Cumming (2002) developed generalized linear models to predict tick distribution in Africa and found that monthly climate variables (temperature extremes, precipitation) added more information than monthly mean NDVI values, while mean annual NDVI performed better than annual climate variables. He suggested that this was related to the temporal lag in the response of NDVI to climatic changes, rendering NDVI a less sensitive variable at finer temporal resolutions (Cumming, 2002).

Where there is obvious and strong congruence between remote sensing-derived land cover classes and biological distributions, retaining sufficient examples of each land cover class has been used successfully as a conservation strategy (Ferrier, 2002). Species richness has also been correlated with habitat heterogeneity, typically represented by an index describing how many different land cover classes occur within a unit (Stoms and Estes, 1993), although elevation range within a transect has also been used (Tognelli and Kelt, 2004). While extremely heterogeneous landscapes tend to have low primary productivity, variation in NDVI can be used to represent heterogeneity (Oindo and Skidmore, 2002). These authors used the standard deviation of interannual (monthly) maximum NDVI values as an index of variation in vegetation structure and composition.

As the functional gradients they represent are location-specific, relationships between species distributions and spectral reflectance values are often not translatable beyond the study area in which the model data were collected. Nagendra (2001) suggests that species–spectral relationships might have to be recalculated for each new image, reducing the overall increase in efficiency of using remotely sensed variables as proxies for biodiversity. Seto \textit{et al.} (2004) suggest that locations outside of their study area may have similar NDVI values but different effects on species richness.

One of the earliest conceptual frameworks for integrating GIS and remote sensing to map species distributions was developed by what is now the US Geological Survey’s National Gap Analysis Programme (GAP; Scott \textit{et al.}, 1993). Originally used to show ‘gaps’ in conservation status of land relative to the distribution of habitat for endangered or threatened vertebrate species, GAP analysis has evolved into a well-established technical and organizational framework for mapping biodiversity ‘elements’ (plant species, communities or habitats) (Jennings, 2000). Jennings (2000) provides an overview of the methods involved: (a) a land-cover map based on Landsat TM imagery is produced by some combination of photogrammetry, supervised and unsupervised classification, and the thematic accuracy is
assessed; (b) land cover maps and GIS data (e.g. elevation, soil) are used in a SDM to produce a habitat suitability map; (c) land stewardship maps are generated, based on biodiversity management categories; and (d) biodiversity elements that are underrepresented in conservation areas are identified. These GAP vegetation and habitat suitability maps produced have subsequently been used as predictor variables for other studies (Frescino et al., 2001). One incidental but no less important result of the success of GAP analysis has been the increase in GIS and remote sensing capabilities among biologists and ecologists (Jennings, 2000).

10.6 Change detection

Land-cover change detection, one of the most common uses of remotely sensed data, is an essential component of ecological monitoring (Aplin, 2005). Change detection and mapping requires land cover maps from at least two time periods (see Figure 10.1), and is possible only if changes in the surface phenomena of interest result in detectable differences in image radiance or emittance (Lunetta et al., 2002). The level of mapping ranges from simple (i.e. change/no change) to complex (i.e. several ordinal change categories) as a function of the dominant land use, prospective disturbance types, management practices and study objectives (Rogan et al., 2003).

The latter scenario is of particular interest to researchers involved in large-area habitat monitoring programmes, where many different types of land cover changes can occur and must be accounted for, e.g. pest infestation, logging and wildfire (Rogan and Miller, 2006). While the most common method of habitat monitoring requires the categorical comparison of independently classified maps, this approach has several drawbacks: (a) high cost (and time consumption) of mapping and re-mapping; (b) inability to detect subtle land cover modifications; and (c) categorical and positional errors in both land cover maps are compounded when compared. The production of maps depicting change can facilitate an improved understanding of both the agents of change and the biophysical linkages between surface reflectance and the change agents, e.g. NDVI can be directly linked to multitemporal changes in green vegetation cover (Lunetta et al., 2002). Rogan et al. (2003) show the increase in accuracy of change maps when GIS variables representing topography are included in the analysis.

10.6.1 Case study: using GIS and remote sensing for large-area change detection and efficient map updating

This case study presents a hybrid change detection technique that integrates GIS and remotely sensed data for efficient change map updating. Two sophisticated, parametric and non-parametric classification techniques, generalized linear models
10.6.1.1 Study area

San Diego County (Figure 10.2) is composed of a variety of heterogeneous land cover types, including shrub-grassland (60%), conifer and hardwood forest (12%), agriculture (6%) and urban (18%). (USFS, 2001). Mean annual precipitation is low (600 mm) and is correlated with elevation, which ranges from sea level to 1991 m. The area is currently undergoing dramatic population growth and accelerated and extensive land cover change due to natural and anthropogenic disturbance (Stephenson and Calcarone, 1999). These spatially and temporally diverse disturbances result in land cover changes ranging from dramatic (e.g. wildfire burn scars, land development) to very subtle (e.g. conifer pest infestation, post-fire regeneration) (Rogan et al., 2002). More than 50% of the county is ‘Category 4’ GAP land management status (i.e. unprotected habitat) (Scott et al., 1993).

10.6.1.2 Data and methods

Two Landsat TM-5 images acquired 8 and 17 June 1996, and two Landsat ETM-7 images acquired 11 and 12 June 2000 (path 40/row 37 and path 39/row 37,
respectively) provided coverage for San Diego County. The images were geometrically registered, resampled and Normalized for atmospheric and illumination effects, using a procedure described in detail in Rogan et al. (2003). Six spectral variables (Kauth Thomas) and five GIS variables (slope, elevation, aspect, vegetation type and previous fire) were used to predict land cover change/no change. The Kauth Thomas variables were: change in brightness (MKT1), change in greenness (MKT2), change in wetness (MKT3), stable brightness (MKT4), stable greenness (MKT5) and stable wetness (MKT6). The modelling dataset was divided into 665 training cases used to develop the models, and 165 test cases to assess the classification accuracy of the model predictions. A full description of the field data collection protocol is provided in Rogan et al. (2003).

Classification trees (CTs) and generalized linear models (GLMs) were used to develop models that predicted probability or suitability of land cover change. The CT model was pruned to 11 terminal nodes, based on cross-validation (Breiman et al., 1984). The GLMs were developed based on a combination of stepwise and subjective, iterative variable addition and subtraction methods, with a goal of minimizing the Akaike information criterion (AIC) statistic (Akaike, 1973; Hastie et al., 2001), using all significant variables. The first GLM was based on linear relationships between change and the predictor variables (GLM_linear). A more complex GLM (GLM_poly) was tested that contained interaction terms suggested by the CT model structure, and non-linear relationships between the predictor variables and likelihood of change. More flexible non-linear relationships were explored using generalized additive models (GAM) (Franklin, 1998; Miller, 2005) and, where appropriate, were specified as polynomials (up to third order) or piecewise linear terms.

10.6.1.3 Results

CT models consist of a series of hierarchical binary splits, the thresholds of which are selected to maximize homogeneity in the two resulting splits. Therefore, the order in which variables are used can be an indication of their relative importance, although scale is also a factor (broad-scale variables are generally used before fine-scale variables). Figure 10.3 shows the CT models for training and test data. Rectangles represent terminal nodes with the majority class (1 = change, 0 = no change), and the fraction below is the number of cases misclassified/total cases (see Rogan et al., 2003).

MKT2 (change in greenness) was the lead split, followed by vegetation, MKT3, MKT6, slope, fire and elevation. The right branch of the tree is associated with increases in the magnitude of greenness values between the two image acquisition dates. It is interesting to observe that this greenness branch was also associated with MKT6 (stable wetness). Due to acute changes in average precipitation levels between the image acquisition dates, soil–plant moisture content likely changed substantially. Recent research has demonstrated the utility of including stable
Figure 10.3 Pruned classification tree models for training (A) and test (B) data. Rectangles show terminal nodes with the majority classification (1, change; 0, no change). Fractions below the terminal nodes give misclassification rate (number incorrectly classified/total number of observations that occur at that node).
features with change features in change mapping, because stable features permit
discrimination among more subtle change classes (Cohen and Fiorella, 1998; Rogan
et al., 2003). This is because different land cover changes can produce similar
spectral signatures in measurement space (Cohen and Fiorella, 1998).

The left branch of the tree is associated with landscape changes related to subur-
bahnization, land clearing and wildfires. These changes can be manifested spectrally,
and produce substantial disturbance-specific changes in soil brightness, green vege-
tation cover and soil–plant moisture content. The most important spectral variables
associated with disturbance were MKT2 (change in greenness), MKT (change in
greenness) and MKT6 (stable wetness). These changes are also correlated with
topographic variables. Gradual slopes (<4.7°) were associated with change (most
likely development), as were low elevations (<104 m). This evidence compliments
the association between changes in greenness, slope and elevation, and reveals a
scenario of decreases in green vegetation cover and soil–plant moisture at low
elevation levels and shallow slope gradients in San Diego County for the time
period examined. While such a scenario is not necessarily surprising to local ecolog-
ic experts and landscape planners, the results of this case study provide spatially
explicit information on the actual locations of landscape changes in association
with myriad disturbance agents and events. Currently, this information can only
be revealed in this spatially continuous context at a 30 m minimum mapping unit,
using the integration of remotely sensed data, environmental variables and statistical
models in a GIS.

Table 10.3 illustrates how the ability to specify non-linear relationships improves
the GLM’s ability to predict probability of change. Figure 10.4 shows the general-
ized additive model (GAM) plots of the smoothed effect of the predictor variables
on change. Elevation, MKT1, MKT2 and MKT3 all appear to have non-linear
effects on probability of change. When specified as polynomials in the GLM,
all four variables except MKT3 improved the model fit (Table 10.3). MKT2 is
also the most important variable in GLM_poly, and was the most important non-
categorical variable in GLM_linear. GLM_poly has a lower AIC statistic, indicating
that it is a parsimonious model that fits the data much better than GLM_linear.

<table>
<thead>
<tr>
<th>Table 10.3</th>
<th>Model AIC and reduction in deviance for each variable used in GLM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>Fire</td>
</tr>
<tr>
<td>GLM_linear</td>
<td>53</td>
</tr>
<tr>
<td>AIC = 419</td>
<td></td>
</tr>
<tr>
<td>GLM_poly</td>
<td>53</td>
</tr>
<tr>
<td>AIC = 223</td>
<td></td>
</tr>
</tbody>
</table>

Numbers in parentheses show polynomial order.
*Variable was insignificant at p < 0.01.
Figure 10.4  Response curves based on GAMs for predictor variables that were specified to have non-linear relationships with change.

Model accuracy was assessed using test data and the area under the curve (AUC) of receiver-operating characteristic (ROC) plots. ROC plots are a threshold- and prevalence-independent metric used to measure how well a model can discriminate between two outcomes: AUC = 0.5 shows no ability, while AUC = 1.0 shows perfect ability (Fielding and Bell, 1997). The non-linear GLM also had higher model accuracy than the linear GLMs. (Figure 10.5).

The GAMs in Figure 10.4 show the non-linear relationships between probability of change and elevation, MKT1, MKT2, MKT3 and MKT4. Elevation is strongly inversely related to change in San Diego County, as the urban development occurs in the lower elevation area near the coast. MKT variables are expected to have a non-monotonic relationship with change, as they show increased values (> zero).
Figure 10.5 ROC plots for all three models show that the more flexible models (GLM_poly, CT) have higher accuracy than the linear GLM model and decreased values (< zero) in brightness, greenness and wetness. The GAM plot of elevation shows a distinct threshold at 400 m, above which change is not likely to occur, and below which probability of change increases as elevation decreases. This particular threshold does not appear in the classification tree (although the threshold of 104 m is used in a low split), most likely because elevation is correlated with many of the MKT variables, particularly change in greenness (MKT2), and this effect is captured when these variables are used.

Classification accuracy based on the training data provides an optimistic view of model accuracy: AUC for the CT and GLM_poly models were both high (0.97 and 0.96, respectively) and not significantly different. Classification accuracy using the
test data was also high for the CT and GLM_poly models (AUC = 0.93 and 0.94, respectively). Accuracy for the GLM_linear model was lower and significantly different for both training and test data (AUC = 0.85 and 0.80, respectively).

10.6.1.4 Case study discussion

A comparison of state-of-the-art mapping algorithms was performed to detect land cover change in San Diego County, California. Non-linear relationships between the spectral and GIS variable to change/no change were important, as was the combination of GIS and remotely sensed predictor variables. The success of the non-linear methods is related to the evidence (from GAMs and linear GLMs) that land cover modifications are complex and have multiple causative effects, and are therefore not amenable to traditional linear modelling approaches.

10.7 Conclusions

An understanding of the environmental factors that determine species distributions (type, abundance, level of diversity) has always been of great interest in ecological research, but its importance has increased along with interest in studying the consequences of changing environmental conditions. The ability to map, model and monitor these distributions is dependent upon the ability to collect, manage and analyse data that adequately describe them. GIS and remote sensing have become indispensable tools in this regard, providing increasingly more ecologically relevant data at higher spatial and temporal resolution, as well as the methods to derive more information from the data and to analyse them statistically.

A more extensive integration of GIS and remote sensing for ecological mapping and monitoring has yet to be fully realized. Much progress has been made in SDM based on GIS data and methods, but investigation into the potential predictive ability of remotely sensed variables, beyond land cover and NDVI, has lagged behind. We see the following as areas in which integration of GIS and remote sensing can have significant effects on ecological research:

- Use of continuous or gradual properties of spectral data should be high priority in SDM; functional gradients can enhance models based on primary and direct gradients.
- More attention should be placed on the ability to consistently characterize biological diversity at multiple scales, using remotely sensed and GIS data.
- In addition to using the reflectance values, the unique synoptic perspective of remote sensing for quantifying landscape characteristics should also be more fully studied.
• Change detection models should explore more flexible statistical methods and include ancillary GIS data.

Acknowledgements

The authors would like to thank Victor Mesev, Ben Gilmer (WVU) and Timothy Currie (Clark University) for reviews and assistance with this manuscript. Funding from the National Science Foundation is gratefully acknowledged.

References


REFERENCES


REFERENCES


