AN ECOLOGICAL PERSPECTIVE ON BIODIVERSITY INVESTIGATIONS: EXAMPLES FROM AUSTRALIAN EUCALYPT FORESTS[1]

Mike P. Austin[2]

ABSTRACT

Australia is a large continent with a relatively small population, and government agencies and research institutions are devoting considerable resources to the development of new approaches and tools for conserving and managing Australia's biodiversity. Issues of data quality, choice of analysis method, ecological theory, and GIS (geographic information system) use are discussed using examples from recent Australian studies with emphasis on the scientific components. The problem of data quality is examined in terms of a suitable minimum data set and the need for a survey design for representative sampling using results from a survey of 24,000 km^2 in northeastern New South Wales. Examples of analytical tools for modeling species distribution, e.g., generalized linear models (GLM) and generalized additive models (GAM), are presented using data from a database of 9537 plots and 273 tree species for an area of 40,000 km^2 in southeastern New South Wales. The necessity for ecological theory, in particular continuum theory as opposed to community concepts, is examined in the context of these results. The interface between ecological and evolutionary theory is discussed drawing on the results of statistical modeling (GLM) of species richness patterns of *Eucalyptus* subgenera in the same area. The predictive use of GIS in mapping vegetation, using statistical modeling (GAM) and multivariate classification techniques, is demonstrated with an application to a comprehensive regional assessment (CRA) process for establishing a regional conservation plan. These methods and analytical tools have been collated into a package, BioRap, which also includes methods for the selection of priority areas for conservation. Rapid progress is being made in developing new tools. However, theory for ecological, statistical, environmental, and evolutionary processes is urgently needed to ensure effective use of these emerging tools for investigating and managing biodiversity.

A key issue facing society is how to conserve our global biodiversity. There is need to use the currently available information now in order to fill the gaps in our conservation strategies. Areas with complementary suites of species and/or representative types of ecosystems are required. There is also a need to constantly examine how to make better use of available data and to find better methods to convert data into useful information for policy decisions on conservation. Scott & Jennings (1998, this issue) presents a detailed account of one of the most comprehensive approaches so far.

Australia is a large continent with a small population, and government agencies and research institutions are devoting considerable resources to the development of new approaches for conserving and managing Australia's biodiversity. To do this, Commonwealth and State governments have developed major databases and Geographic Information Systems (GIS) to provide biodiversity information on the location, abundance, and dynamics of Australia's native flora and fauna, e.g., the Environmental Resource Information Network (ERIN, Chapman & Busby, 1994). Key issues arising from the use of these tools are how best to answer policy questions, data quality, the suitability of analytic tools, the role of ecological theory, the predictive success of GIS, and how best to make methods available to the wider community of users.

This paper focuses on Australian research in this area, in particular on improving information provision methods using modern computer technology. The topics considered are: use of available data, such as herbarium records and vegetation survey data; design of surveys to obtain more cost-effective data; use of statistical modeling and GIS to predict species distributions and richness patterns from survey data; and the need to evaluate methodology against existing theory.

[1] I thank those colleagues in the Division of Wildlife and Ecology and in other Australian organizations whose work has made this review possible. I also thank J. A. Meyers, N. C. Coops, and G. B. Barnett for preparing the figures.
[2] Division of Wildlife and Ecology, Commonwealth Scientific and Industrial Research Organisation, PO Box 84, Lyneham, Canberra, A.C.T. 2602, Australia. *m.austin@dwe.csiro.au*

Data

Most developed countries are establishing databases of biotic data for creating and evaluating conservation policies (Chapman & Busby, 1994; Scott & Jennings, 1998; Soberón et al., 1996). Similar methods are also being adopted for developing countries (Hall, 1994). The data are usually based on herbarium or museum records. In Australia, the federal government has established the Environmental Resources Information Network (ERIN) to collate, organize, and provide access to the available data. Maximizing the use of existing data is now critical as resources to re-collect data by means of surveys are very limited. As part of this effort the principal herbaria and botanic gardens in Australia have cooperated to produce a common standard for computer-based records systems for specimens. There is a working group that meets regularly to address ongoing applications issues. It is estimated to cost $6 Australian to database a single herbarium or museum record, but several times that to collect specimens using professional staff (Chapman & Busby, 1994). ERIN has developed an extensive hardware and software system to support the aim of providing primary data to identify and characterize regional environmental patterns for use in environmental assessment and planning. For handling taxonomic data, ERIN has developed modules for managing taxon names and easily updating them (*Taxon*), managing individual records of specimens *(Specimen)*, and a *Data Dictionary and catalogue* module for managing data sets including custodianship. These modules and others are linked to a GIS to form what ERIN terms a Spatial Information System (SIS). Chapman and Busby (1994) provided further details of the system, and there is a website (http://www.erin.gov.au) that also provides a public access system for plant records. The system provides for all types of data and remote-sensing coverage of Australia, but the primary taxonomic record data are a key component.

It is important, however, to recognize that herbarium records suffer from several weaknesses (Hall, 1994; Margules & Austin, 1994; Soberón et al., 1996): the records record presence only, and there is no information about absence; the locations are often poorly recorded; the presence of other species and of environmental variables is inconsistently recorded; and the spatial distribution of specimens is highly biased. Figure 1 exemplifies the location bias of museum records; it shows the distribution of all suitable records of elapid snakes in Australia (Longmore, 1986). The major roads in remote areas of the continent are clearly outlined by the record locations. With hindsight it is easy to be critical, and such records were not intended to provide definitive data for regional biogeographic or conservation studies. However, when used for analysis of areas of high biodiversity or endemism, problems can arise; see Tuomisto (1998, this issue) for an example from Amazonia. The minimum data set needed for analysis is presence/absence data and an accurate location for which environmental data can be obtained via a map or GIS. Statistical analysis is precluded by the lack of absence data. How to use presence-only data is a serious problem not always recognized by systematists (Soberón et al., 1996; Margules & Redhead, 1995) when considering conservation issues. However, herbarium collections provide taxonomic precision and verifiable voucher specimens, which vegetation surveys usually lack.

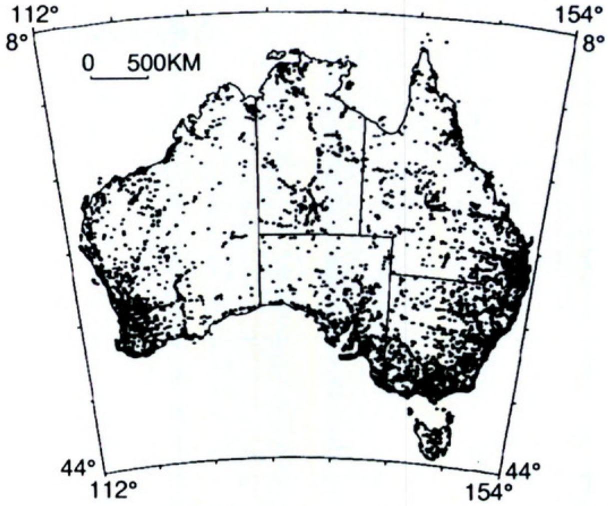

Figure 1. Collection sites for all species in the *Atlas of Elapid Snakes of Australia* (Longmore, 1986). Note the alignment of sites along major roads, especially in the Northern Territory. Reprinted with permission.

This has led to the development in Australia of two heuristic methods to make maximum use of presence data. The first, BIOCLIM (Nix, 1986; Busby, 1986, 1991; now termed BIOMAP (Hutchinson et al., 1997)), uses geocoded specimen records together with estimates of a selected set of bioclimatic variables for the location. The estimates are derived from climatic surfaces calculated using records from climatic stations. These specimen records are used to estimate the range of each bioclimatic variable within which the species is found. For each location where a specimen of a species is recorded, the climatic estimates are aggregated to provide a "climate profile" of the taxon. The values for each estimate are ranked in increasing order such that the minimum value, the 5th percentile, and 95th percentiles, etc., can be defined. This has been done for 12 bioclimatic variables to define the climatic profile (Busby, 1986). By describing the climatic profile for a species as the combination of climatic conditions lying between the 5th and 95th percentiles for 16 climatic variables, a climatic envelope for the *potential* occurrence of a species is defined. From this profile, together with a grid of predictions of the bioclimatic variables for a region or continent, a map of the potential occurrence of a species can be generated based on climatic information (Longmore, 1986; Hutchinson et al., 1997). The prediction map is only of potential occurrence because no information on absence is used, and there is no information on other environmental or historical factors that might control species occurrence.

There are four essential components to the procedure: (1) a method to produce climatic estimates from the records of climate stations and measurements of latitude, longitude, and elevation (see Wahba & Wendelburger, 1980; Hutchinson & Bischof, 1983; Hutchinson, 1984); (2) existence of a digital elevation model which can be used to generate the climatic predictions for all points in the region; (3) a conceptual model for deciding on an appropriate set of bioclimatic variables relevant to the organisms being studied (Nix, 1986); and (4) a classification algorithm to define the bioclimatic envelope. Hutchinson et al. (1997) provided an up-to-date presentation of all stages of the approach. Examples of the application of this method are Longmore (1986), for a continental study of elapid snakes; Nix and Switzer (1991), on the potential regional distribution of Australia's rainforest vertebrate fauna; Busby (1986) on distribution of *Nothofagus cunninghamiana* (Fagaceae); and Busby (1988) on the impact of climate change on Australia's flora and fauna.

A revised method, HABITAT, has been published (Walker & Cocks, 1991) that uses a polygonal rather than the cruder multidimensional rectangular definition of the climatic envelope used by BIOCLIM. This procedure provides a more conservative (smaller) envelope that takes more account of the actual distribution of presence records in the climate space. It has been applied to estimating the continental distribution of kangaroos (Walker & Cocks, 1991). However, BIOCLIM (now BIOMAP) remains the most extensively used of heuristic methods for presence data. See Austin et al. (1994a) for a further review of presence methods.

To provide better data, herbarium records should contain precise locations and consistent environmental information. A preferable minimum data set is presence/absence data for all species in a standard set of taxa from plots collected as part of a vegetation survey. In any survey, absence is conditional on the sampling effort made at a site. Large databases that are capable of supporting statistical modeling can be built up by collating such data from existing surveys (Austin et al., 1990; Leathwick & Mitchell, 1992). The principal weakness in such data sets is the unknown sampling bias in the original selection of the plots. To make most use of databases they should be ecological in nature rather than taxonomic. Margules and Austin (1994) have discussed the requirements for establishing such a database, listing four requirements: (a) a conceptual framework based on ecological theory; (b) field data obtained from sites using survey design principles based explicitly on the conceptual framework; (c) a rationale for determining which measurements should be made at the chosen sites in addition to the floristic records; and (d) appropriate statistical methods for analyzing survey data and predicting (extrapolating) the regional distribution of species from the point records. These authors failed to emphasize that this is only possible if the database is linked to a GIS.

Survey Design

How to obtain a representative sample of the vegetation variation in a region is a central question for conservation evaluation. Vegetation surveys of large areas are expensive and time-consuming, particularly if random or systematic sampling is undertaken in rugged or inaccessible regions (Burbidge, 1991; see also Tuomisto, 1998). Cost-effective methods are required. In Australia, modifications by Austin and Heyligers (1989, 1991) of the gradsect sampling approach first proposed by Gillison and Brewer (1985) provide an example of an explicit, consistent, and repeatable method. Unlike many sampling strategies that produce unbiased estimates of some mean value, e.g., basal area of timber per unit area in the region, vegetation surveys should be directed toward obtaining a representative sample of the **range of variation** in vegetation composition. The detection of unusual combinations of species is as important as accurate estimates of the average composition of the commonest forest types. The method proposed by Austin and Heyligers (1989) is based on sampling vegetation from all possible combinations of selected environmental variables. The logistics of surveys, e.g., travel time between sampling sites, add considerably to the costs. Sampling along a transect is very cost-effective in travel time. If the transect is oriented along the steepest environmental gradient in

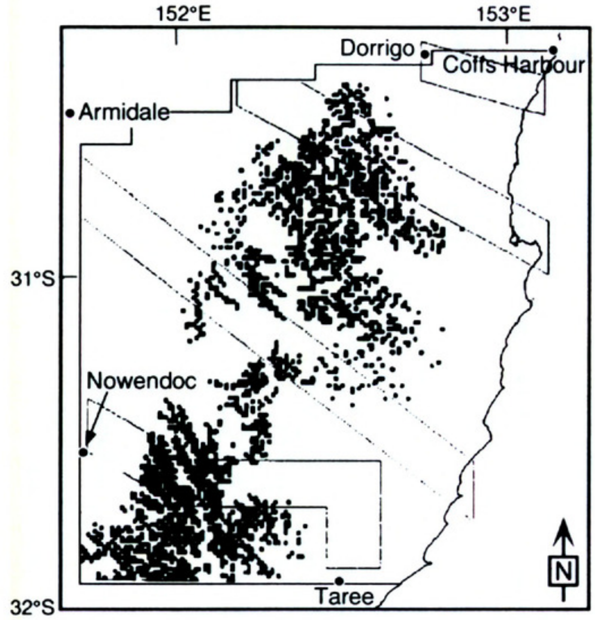

Figure 2. The position of four gradsects selected for a region of the north coast of New South Wales, showing the extent to which they sample a particular altitude/rainfall class (mean annual rainfall 1000–1399 mm and altitude 180–540 m). Individual squares represent 1 km². Redrawn from Austin and Heyligers (1991).

a region (i.e., a gradsect), then different environments can be sampled with less effort. Where such a gradsect is positioned along an access route, then a very cost-effective although biased survey is obtained.

Austin and Heyligers (1989, 1991) designed a survey of the forest vegetation of coastal northern New South Wales (NSW) based on the principles outlined above. The area surveyed was 24,000 km², and the floristic data consisted of presence/absence data of tree species recorded from a 50 × 20 m plot oriented along the contour with estimates of the ranking of the dominant species. The protocol used consisted of seven steps: (1) Identify the major environmental variables influencing the distribution patterns of the vegetation in the study region. For their region these were temperature, rainfall, radiation, and nutrients. (2) Recognize a set of variables best suited because of their availability and practicality to determine the position and direction of the gradsects. For the north coast region these were altitude (an easily measured and highly correlated surrogate for temperature), mean annual rainfall, and lithology (crude surrogate for soil nutrient content). (3) Select gradsects using these variables and the best available technology. Figure 2 shows the extent to which the four selected gradsects sample one particular combination of altitude and rainfall. (4) Stratify the gradsects into geographical segments and stratify the environment within segments to provide

Table 1. Example of survey design for a segment. Size of the environmental cells and their sampling frequency for the middle segment of the southern segment. Reproduced from Austin & Heyligers (1989).

Rock-type	Altitude classes	Rainfall classes 2	3	4	5	6
	1					
	2				4 (x)	
3	3				3 (1)	45 (1)
	4					2 (x)
	5					
	1					
	2					
6	3			1 (0)		
	4			1 (0)		
	5					
	1	2 (1)	15 (1)	26 (1)		
	2		16 (1)	37 (1)	38 (1)	1 (0)
8	3		1 (0)	34 (1)	22 (1)	12 (1)
	4					1 (0)
	5					
	1		11 (1)	30 (1)	2 (x)	
	2		55 (2)	17 (1)	28 (1)	1 (0)
9	3		86 (2)	75 (2)	4 (x)	2 (1)
	4		1 (0)	26 (1)		
	5			1 (0)		

The plain numbers refer to the total number of pixels in an environmental cell. The numbers in parentheses refer to the number of samples (a sample may consist of up to 5 plots each from different topographic positions) to be sampled in each cell. An "x" indicates that these cells were not easily accessible and no sample was to be taken.

replicate sampling of different environmental combinations at different locations (Fig. 2). (5) Stratify at the local scale, i.e., within the 1-km resolution used in positioning the gradsects to take account of other important environmental determinants of vegetation. In this case five topographic positions as a surrogate for solar radiation were sampled within each 1-km gridcell selected. (6) Decide the effort to be spent sampling the rarest environmental combinations as compared with increased replication of the commonest combinations. Determine the location of samples by selecting random coordinates and taking the closest suitable cell with adequate access. Table 1 shows an example of the survey design for a particular segment. (7) Review assumptions regarding importance of environmental variables on which the survey was designed and modify if necessary before completing the survey. Austin and Heyligers (1989) modified the survey design after finding that depth to water table had an overriding influence in the coastal lowlands.

The approach of Austin and Heyligers (1989) can be summarized as an **SR**³ strategy, that is, **S**tratification, **R**epresentation, **R**eplication, and **R**andomization. It represents one realization of the first two requirements of Margules and Austin (1994) for establishing a database. A total of 1025 plots were sampled by Austin and Heyligers (1989) equal to an area of 1.025 km^2, or approximately 1/24,000th of the study area. The restriction of sampling to gradsects means that not all locations have an equal chance of being selected in the sample, and therefore the sample obtained is highly biased. However, the sample is representative, and the design is explicit, consistent, and repeatable, which is not always the case with biodiversity surveys at the present time. While it is possible to design an SR^3 survey without a GIS, it is much easier if one is available. Modifications of it have since been used in northern NSW, Australia (Ferrier, pers. comm.), in Sri Lanka (Green & Gunawardena, 1993), and in South Africa (Wessels et al., in press). The approach makes two ecological assumptions. First, that vegetation varies continuously with environment, forming a continuum rather than discrete communities (Austin & Smith, 1989), and therefore all combinations of environmental conditions should be sampled. Second, it assumes that the major environmental gradients in any given region are known or can reasonably be hypothesized.

The comparative performance of gradsect sampling has been evaluated by Austin and Cawsey (1991) with artificial data and by Wessels et al. (in press) with survey data for birds and dung beetles in South Africa. Both studies support the cost-effectiveness of gradsect sampling against systematic, random, and various purposive methods. Detailed attention has been given to this example of survey design because vegetation survey methods appears to be a neglected topic (Greig-Smith, 1983; Jongman et al., 1987; Kent & Coker, 1992), though see Noy-Meir (1971) for an early Australian example. A variety of alternative designs have been developed in Australia; see Noy-Meir (1971), Austin and Basinski (1979), Margules and Nicholls (1987), Prober and Austin (1990), McKenzie et al. (1991), and Neave et al. (1996).

Statistical Modeling

Availability of survey data consisting of presence/absence for species plus information on environmental variables from a GIS allows the prediction of species distributions using statistical modeling. Statistical modeling is no longer restricted to quantitative data with normal errors (McCullagh & Nelder, 1989), as many botanists assume. There is currently a wide variety of prediction methods extending well beyond the usual statistical methods to neural nets (Aleksander & Morton, 1990; Fitzgerald & Lees, 1992), genetic algorithms (Holland, 1992; Lees, 1994), and decision trees (Breiman et al., 1984; Lees & Ritman, 1991). A recent evaluation of many of these methods (Austin et al., 1994a, 1995; Austin & Meyers, 1996) for analyzing plant ecological data concluded that while most techniques can be found to have advantages under certain circumstances, statistical models perform better with typical vegetation survey data. Franklin (1995) provided review of recent work from a geographer's perspective. The two statistical modeling methods that are currently being actively used are Generalized Linear Models (GLM; McCullagh & Nelder, 1989) and Generalized Additive Models (GAM; Hastie & Tibshirani, 1990). Examples of the use of GLM with vegetation data are Austin et al. (1990, 1994b) and Leathwick and Mitchell (1992). Nicholls (1989, 1991) provided a detailed discussion with examples of how to use GLM with vegetation survey data. The more recent technique of GAM was introduced to plant ecology by Yee and Mitchell (1991). Leathwick (1995) used it to study the climatic relationships of New Zealand tree species. Austin and Meyers (1996) compared GLM and GAM for *Eucalyptus* forest species and discussed their role in the management of forest biodiversity. Recently GAMs have been used for predicting flora and fauna distributions for a large area of northwestern NSW (NSW NPWS, 1994a, b) and to derive predicted vegetation communities for the south coast of NSW in an unpublished CSIRO consultancy report in 1996.

Statistical models such as GLM are used for the prediction of a response variable (or dependent variable) from a set of predictor (or independent) variables. One advantage of GLM over the classical regression method is that it allows error functions other than the normal, and hence the use of density or even binary data is possible. GAM, a non-parametric technique, has the additional advantage that the mathematical function describing the shape of the curve relating the response variable to a predictor variable need not be specified precisely. A smoothing spline is fitted to the data, and only the number of inflections in the curve need be specified, not whether it is a polynomial or exponential function.

The key problem in the model-building process for GLM use with vegetation data has been the shape of the response of a plant species to environmental predictors. Ecological theory is needed to define a reasonable set of potential responses. The evidence regarding the existence of the bell-shaped response usually presented in textbooks is ambiguous (Austin & Smith, 1989), and more flexible curves need to be considered. The β-function is one complex function that has been proposed (Austin et al., 1994b). It requires definition of the limits of a species distribution along an environmental gradient within which a variety of skewed or symmetric curves can be represented by β-functions with different parameter values. Austin et al. (1994b) fitted a β-function for temperature to data for nine species of *Eucalyptus* (Myrtaceae). No species had a symmetric response shape; all were skewed and the patterns of skewness were dependent on position along the environmental gradient of mean annual temperature (Austin et al., 1994b). The results were confirmed for a larger set of *Eucalyptus* species (Austin & Gaywood, 1994). Their conclusions suggest that species distributions along gradients have well-specified skewed shapes and nonrandom patterns. If these patterns are found in other suitable data sets, then it may be possible to propose rules regarding the biodiversity patterns to be found in vegetation. Data sets are needed where the length of the environmental gradient sampled clearly exceeds the width of the environmental niche of the individual species, otherwise the species limits cannot be specified. Failure to appreciate this limitation to the use of β-functions has resulted in controversy (Oksanen, 1997; Austin & Nicholls, 1997; see also Austin & Meyers, 1996).

It is the difficulty of specifying the exact form of the response shape that has led several researchers to use GAM (Yee & Mitchell, 1991; Norton & Mitchell, 1993; NSW NPWS, 1994a, 1994b; Leathwick, 1995). GAM, while conferring the advantage of a non-parametric smoothing function, is not without problems, e.g., the sensitivity of significance tests (Austin et al., 1995; Austin & Meyers, 1996), and is certainly not without assumptions as asserted by Norton and Mitchell (1993). It is a "current best practice method" for biodiversity analysis but is likely to undergo significant modifications in the near future as further evaluation is done.

Eucalyptus fastigata: A Case Study in GAM Prediction of Species Distribution

The steps in modeling the distribution of a species based on available plot data from a region in southeastern NSW, Australia, are presented here. The details of the study area have been published previously (Austin et al., 1990; Austin et al., 1994b). Briefly, it is approximately 40,000 km^2 in area and runs from just north of latitude 35°S to the Victorian border, and from longitude 148°E to the east coast of Australia. The climate varies considerably across the region: mean annual temperature ranges from 2.5°C on Mt. Kosciusko at 2200 m to 16.9°C on the northern coastal plain, while mean annual rainfall varies from 480 mm to more than 2000 mm, with marked seasonal differences in rainfall patterns. *Eucalyptus fastigata* H. Deane & Maiden is a species characteristic of the coastal scarp forests between 400 m and 800 m (Fig. 3), and results of statistical models of its distribution using GLM have been published (Austin, 1992; Austin et al., 1994b).

Modeling Steps

1. Collate available plot data for the defined region. Details of the database and the contributors can be found in Austin et al. (1994b). The current database has 9537 plots with records of the presence/absence for 273 tree species, and the geographical distribution of the plots is shown in Figure 3.
2. Select and generate a set of appropriate environmental predictors. Climatic variables have been generated from a digital elevation model (DEM) using a variety of packages now incorporated into BioRap (Hutchinson et al., 1997), and values for the plots obtained using a GIS. Those variables based on lithology were derived from a lithology GIS layer. Eleven environmental variables were selected as potential predictors. These included eight continuous variables: average summer rainfall, average winter rainfall and rainfall seasonality (ratio of summer/winter rainfall), mean annual temperature, temperature of the hottest month and winter cold index, average summer radiation and average winter radiation (kj/m^2/day adjusted for slope and aspect); and three factor or categorical variables: topographic position (6), lithology class (6), and nutrient index (5) (figure in parentheses is the number of classes in the factor). Figure 4 shows the plots mapped into a climate space equivalent to the geographical space shown in Figure 3.
3. Restrict the data, where the data extend well beyond the environmental niche of a species. For example, *E. fastigata* does not occur below 7°C or above 16°C mean annual temperature (Fig. 4). Inclusion of those zero values beyond these limits can complicate the analysis and lead to poor prediction of the occurrence of the species near the limits (Austin & Meyers, 1996). The data are therefore restricted to those observations that occur within limits set by having

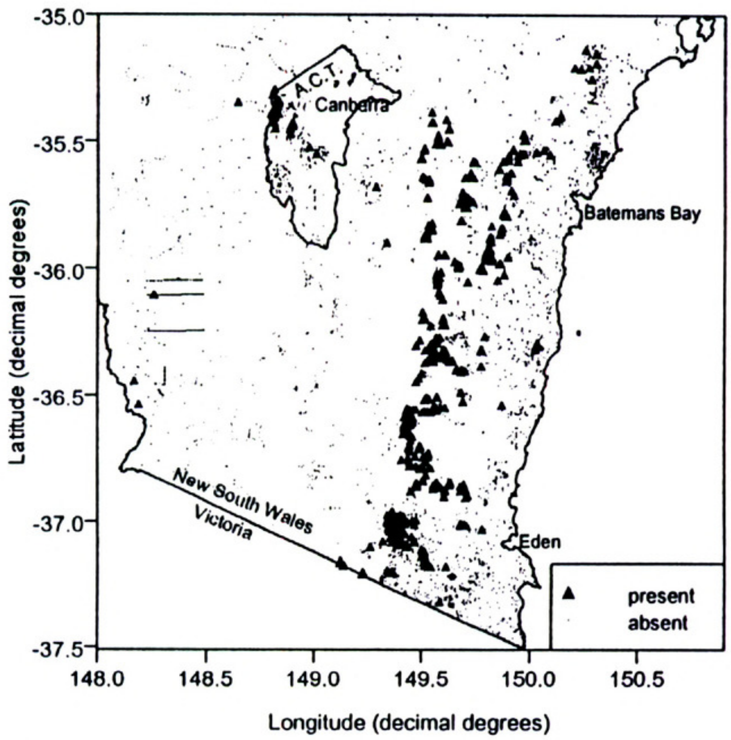

Figure 3. The geographical distribution of *Eucalyptus fastigata* as determined from the database of 9537 plots for southeast NSW. Triangles indicate *E. fastigata* present; dots indicate plots without *E. fastigata*.

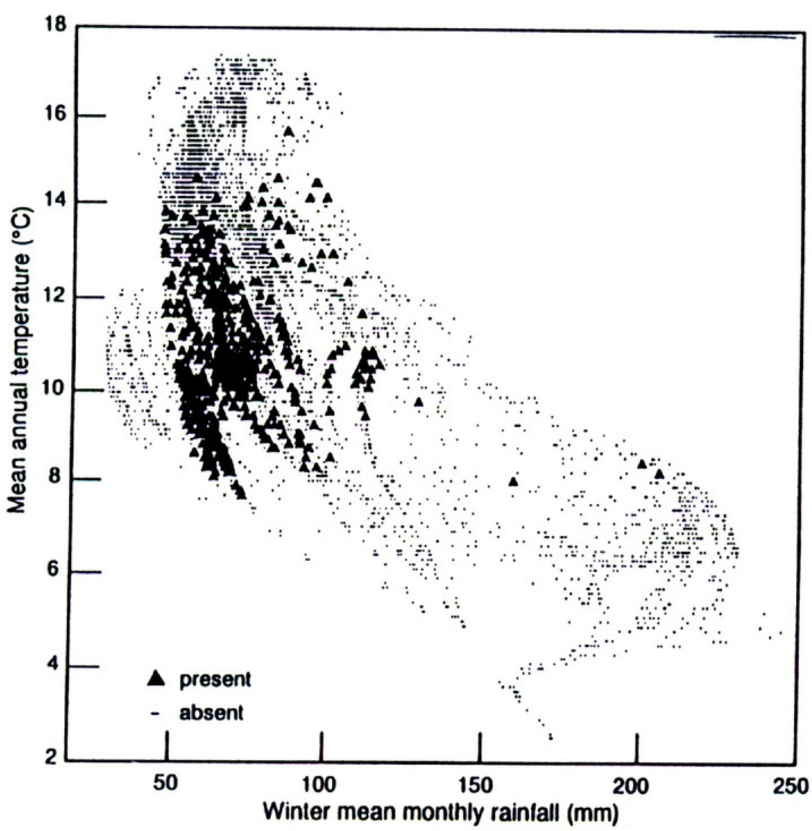

Figure 4. The distribution of *Eucalyptus fastigata* in a climate space defined by mean annual temperature and mean monthly winter rainfall. Note *E. fastigata* is absent below 7° and above 16°C mean annual temperature, with numerous plots above and below those limits.

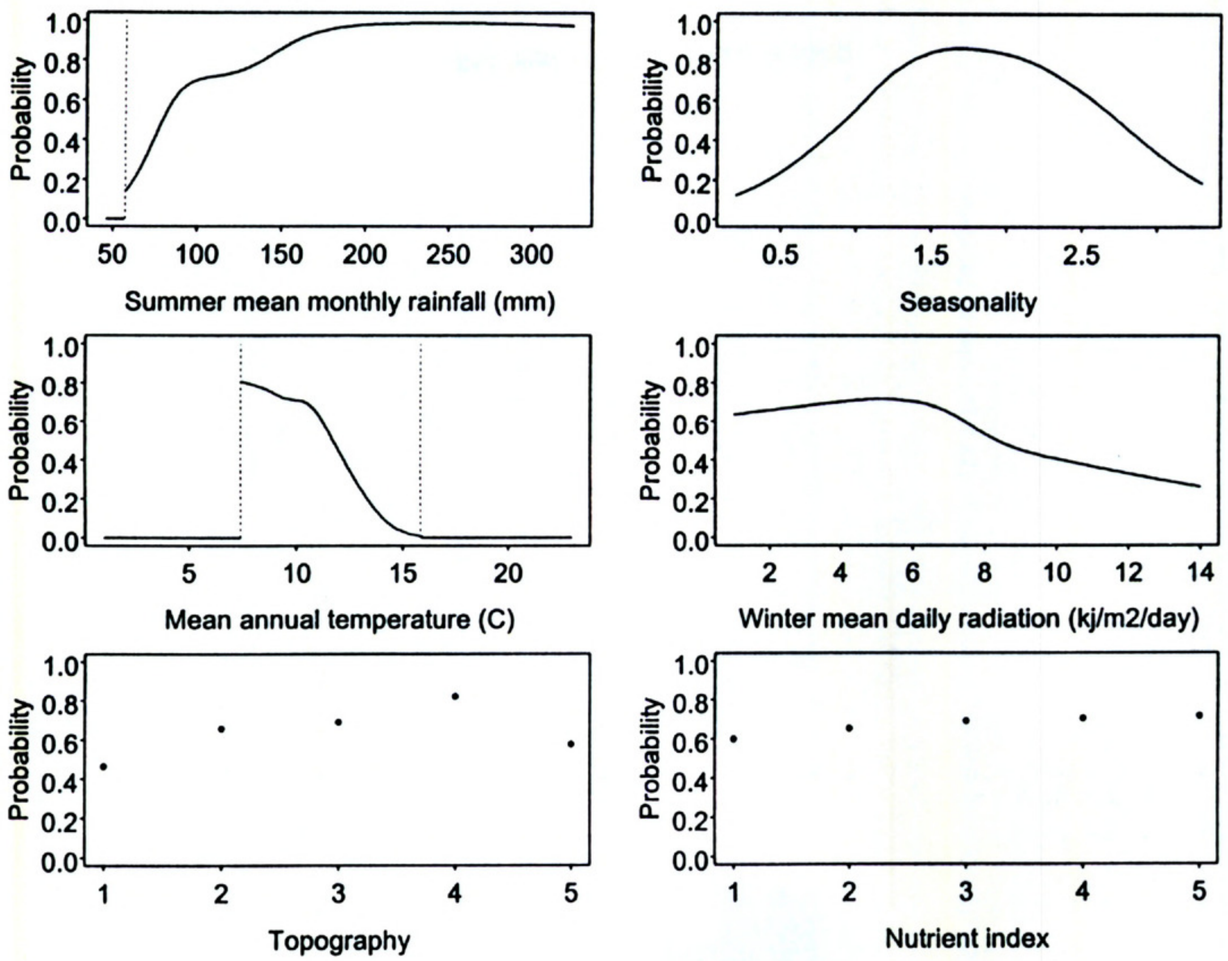

Figure 5. The shape of the GAM response functions for 6 of the 11 predictors for *Eucalyptus fastigata.* Note that the functions have been fitted only within limits for mean annual temperature and summer mean monthly rainfall.

100 zero values above and below the last positive observation, provided there are additional observations beyond the limits; see Austin and Meyers (1996) for further details.

4. Fit a GAM. The model was derived for presence/absence data for *E. fastigata,* as predicted from the 11 environmental variables using S-Plus package (Statistical Sciences, 1993), with four degrees of freedom for the continuous predictors. All eleven predictors were included in the model. The shapes of the responses differ markedly for the different predictors (Fig. 5).
5. Use GIS to predict the distribution of the species for unsampled areas in the region. This was done using the predictive functions derived from GAMs. The predicted distribution of *E. fastigata* clearly shows the major zone of occurrence along the coastal scarp (Fig. 6; cf. Fig. 3).

These models can be used to investigate current ecological problems of relevance to our future management of biodiversity. For example, where would *Eucalyptus fastigata* occur if global warming resulted in a 2°C rise in regional temperature and local increases in rainfall? The predicted geographical distribution after such a change is shown in Figure 7. *Eucalyptus fastigata* would undergo a substantial reduction in occurrence on the coastal scarp under such a scenario. Note that this is a static analysis ignoring problems of dispersal, time to equilibrium, and changed competitive interactions. The role of environmental niche models in relation to climate change models and physiological growth models was reviewed by Austin (1992), with particular reference to *E. fastigata.*

The above procedure is explicit, repeatable, and consistent. There are both statistical and ecological research issues still to be resolved about the best procedure. Austin and Meyers (1996; Austin et al., 1995) have examined the performance of GLM, GAM, and regression trees on both real data and simulated data where truth is known. They conclude that a mixed strategy using both GLM and GAM functions is desirable. They suggest that the best results are as dependent on the availability of suitable ecological and statistical skills as on the particular procedure used. The explicit nature of

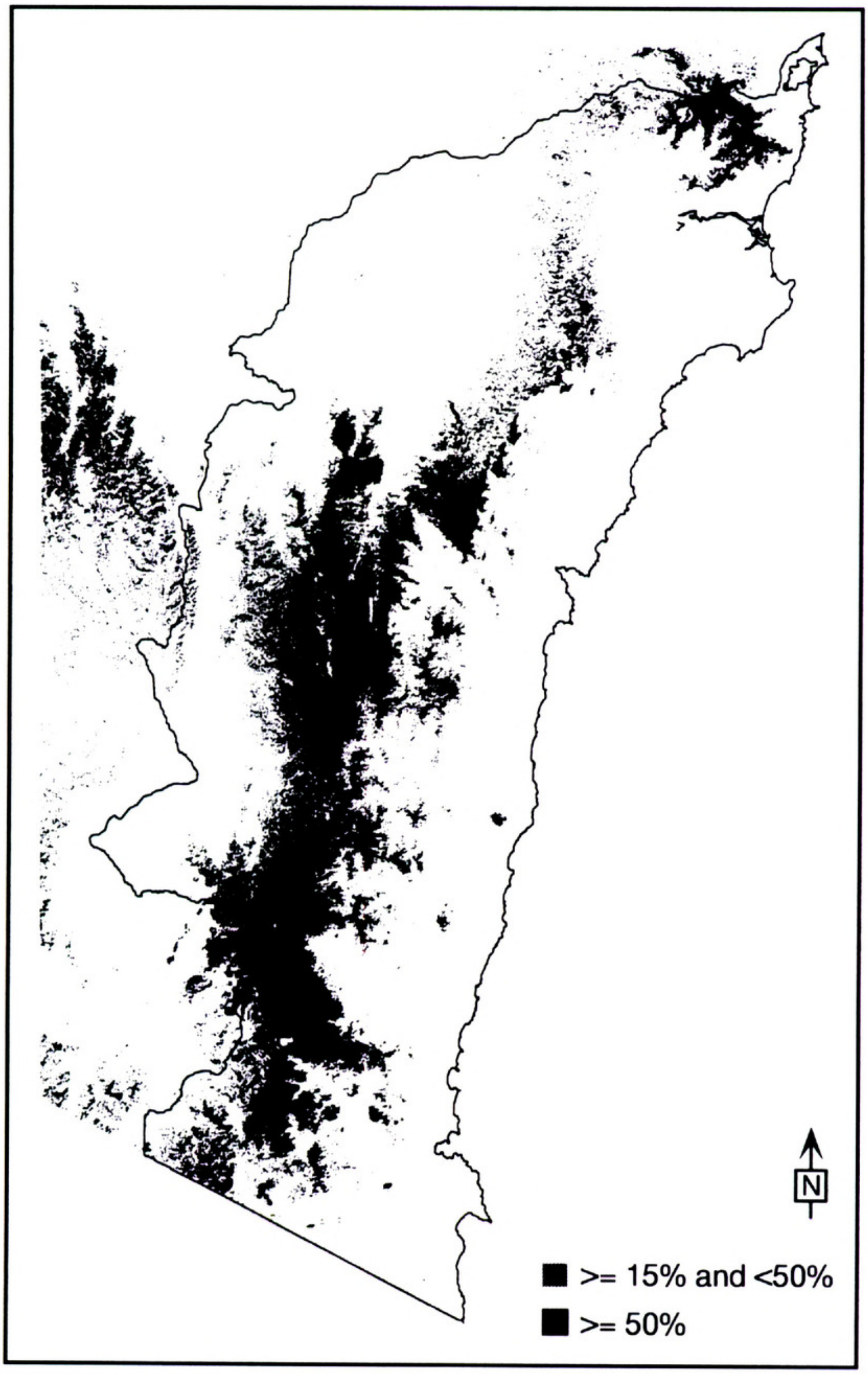

Figure 6. The predicted geographical distribution of *Eucalyptus fastigata* in terms of probability of occurrence using the GAM functions and a GIS for the coastal zone (outlined area) of New South Wales.

the models for individual species provides a firm basis on which to build an improved understanding of species distribution patterns. The *ad hoc* mapping of vegetation and species based on unknown mental models derived from an unknown arbitrary database imperfectly remembered is no longer adequate. However, it must also be remembered that these models are only as good as the data, and the ecological assumptions on which the predictors are selected, and are based on correlation, not causation.

Species Richness

Statistical models like GLM can be used in other contexts more relevant to evolutionary botany and

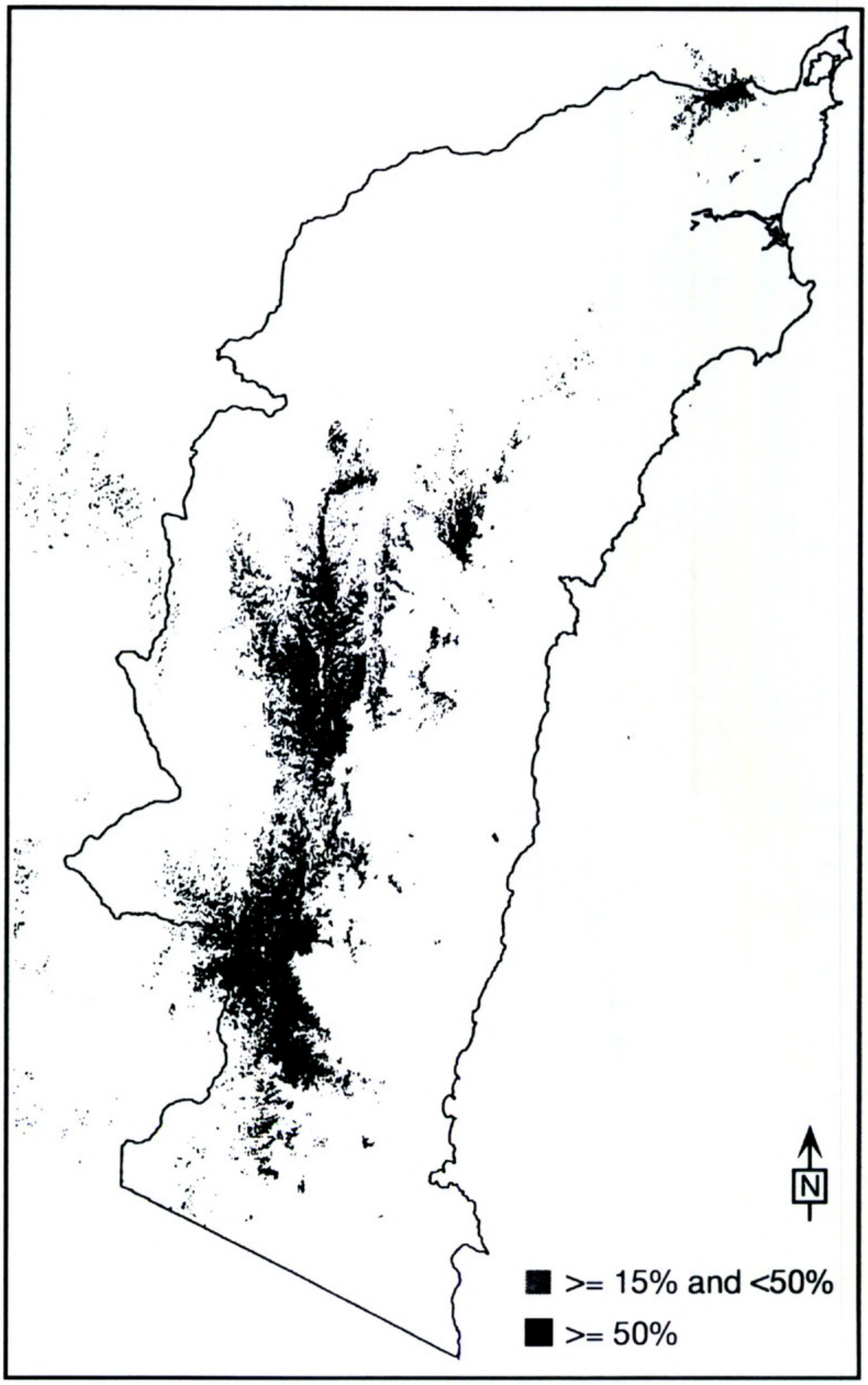

Figure 7. The predicted distribution of *Eucalyptus fastigata* if mean annual temperature increased by two degrees and rainfall by 20% in coastal and tableland regions and 10% in the western region for the same region as Figure 6.

its interface with ecology (Currie, 1991). Austin et al. (1996) investigated patterns of tree species richness in southern NSW using a similar but smaller data set (7208 plots) than that used for *Eucalyptus fastigata* above. Similar predictors to those of Austin et al. (1994b) were used, namely mean annual temperature, mean annual rainfall, mean annual daily radiation, and four categorical variables (topographic position, lithology, nutrient index, and rainfall seasonality). Total tree-species richness for 0.1-ha plots was predicted as the dependent or response variable using GLM, with cubic polynomial functions for the continuous variables and interaction terms for temperature and rainfall. Regional scale patterns of species richness are predictable from the environment, with mean annual temperature the most important predictor. Maximum species richness for trees was found in protected gullies at temperatures >16°C, with rainfall >900 mm, on volcanic soils with intermediate or high nu-

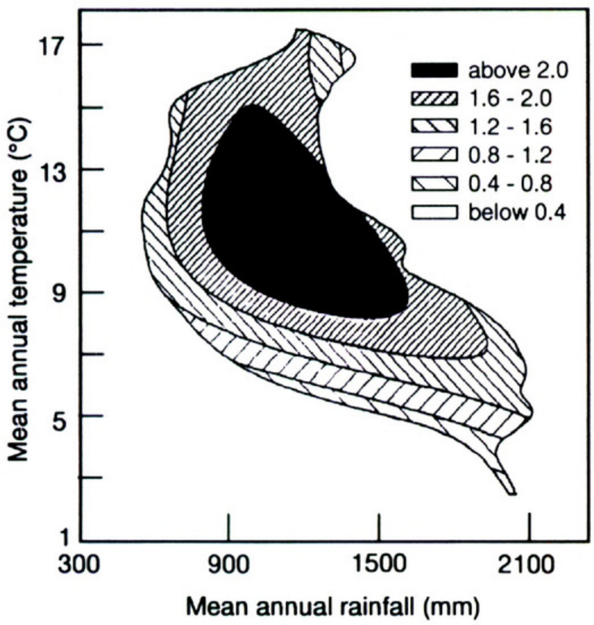

Figure 8. The predicted distribution of species richness for *Eucalyptus* subg. *Monocalyptus* in relation to climatic predictors on exposed ridges with high radiation and low nutrients, with soft sedimentary lithology.

trient levels (Austin et al., 1996). This habitat represents the limited conditions under which the species-rich warm temperate rainforest species can survive in the fire-prone eucalypt-dominated forests of the region. Various components of tree-species richness can be recognized. For example, there are numerous species of *Eucalyptus* in the region, and analyses of the species richness patterns were made for two of the subgenera, *Monocalyptus* and *Symphyomyrtus*. All predictors except seasonality of rainfall were significant for the subgenus *Monocalyptus*, and there was a complex skewed response to temperature and rainfall (Fig. 8). Maximum species richness for *Monocalyptus* was predicted for exposed ridges on sediments or granites under low nutrient conditions in temperate climatic conditions. The other subgenus, *Symphyomyrtus*, showed a distinct complementary pattern and insensitivity to radiation and topographic position, but species richness also varied with seasonality of rainfall. High species richness was associated with fertile soils (Austin et al., 1996). There has been considerable discussion about the differential behavior of species from these two subgenera and their ability to co-occur (Noble, 1989). The descriptive models obtained by GLM analysis are consistent with the conclusions reached in the literature review of Noble (1989).

For the evolutionary botanist these results pose the question of why there are more species of subgenus *Monocalyptus* co-occurring in some environments than others. Figure 8 shows that there is an optimum environment for numbers of species of *Monocalyptus*; any theory of biodiversity and evolution should be able to explain the existence of such a pattern in environmental space. Managers of biodiversity also need to understand the relationship between species richness and environment. If individual species and species richness are both strongly related to environment, then the concept of a regional species pool needs to be re-examined.

Applications

An example of the use to which these computer-based tools are being put in Australia is an unpublished consultancy report by the CSIRO Division of Wildlife and Ecology for the NSW National Parks and Wildlife Service. The objective of the consultancy was to map the pre-European forest vegetation (pre-1750) at the scale of 1 : 100,000, such that the percentage of the pre-European communities still surviving could be estimated. This information would then be used to determine which currently forested areas should be conserved, which logged, and which require further detailed examination. The region concerned was the Southern Coastal Zone of NSW. There was only a limited time available to complete the study. However, the existing database and modeling studies described above, plus an appropriate GIS, provided a suitable basis for undertaking the study using modern methods.

The ecological theory on which the study was based assumed that the vegetation formed a continuum such that the precise composition of the vegetation varied continuously, and communities were a function of the frequency of particular environmental combinations in the landscape (Austin & Smith, 1989). Estimating individual species distributions using GAMs from existing data for relevant environments would allow spatial predictions of distributions for cleared areas. Combining the predictions for individual species for each cell of a GIS gives an expected but continuously varying community composition. This composition data can then be classified using numerical methods to give a consistent description of forest communities for the entire region.

The steps involved practical decisions at each stage. These are dependent on the particular features of each project. These are briefly described here to indicate the types of problems that arise.

1. Data. The existing database consisted of 8377 plots. No data were recorded in the database for the northeastern area of the zone. Additional data were collated from that and other regions

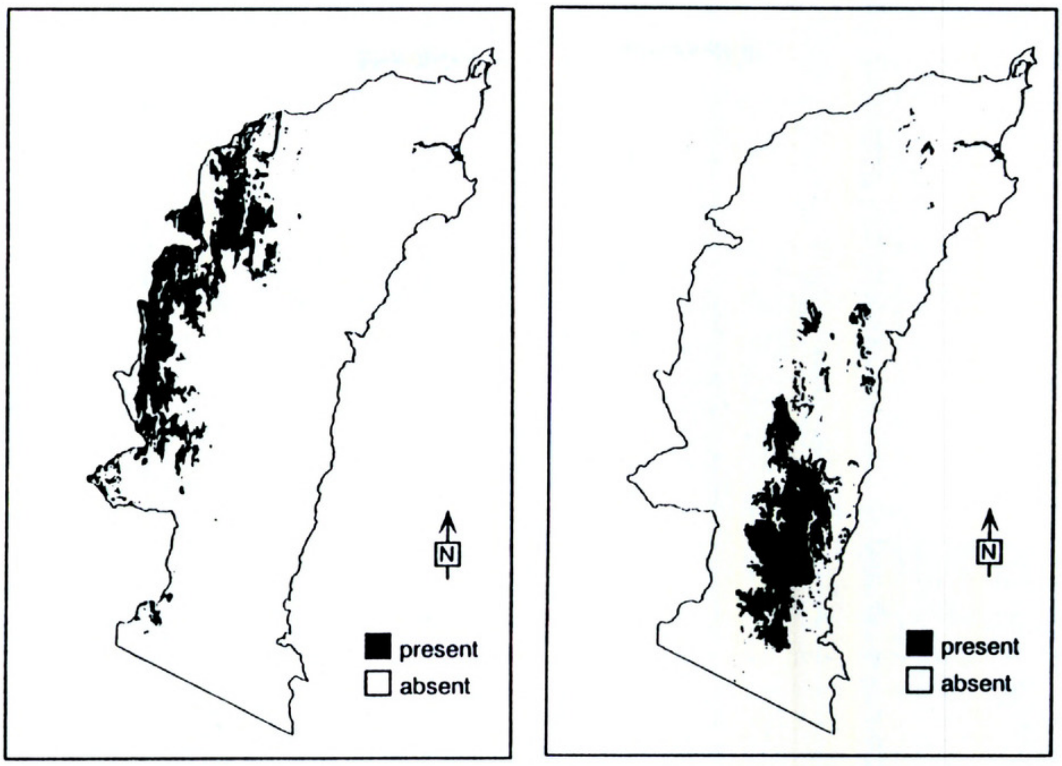

Figure 9. Pre-1750 mapping units. —a (left). Savannah Woodlands (*Eucalyptus melliodora/E. bridgesiana* group). —b (right). Lowland Granite Communities (*E. tereticornis* group).

in the zone. However, this was found to give a very biased sample from the northeastern region; most plots were from warm temperate rainforest in gullies. An additional survey of the northeastern region was necessary.

2. Survey. The design was based on the **SR**³ strategy, described previously. The gradsect approach was not used, as access was not a major limiting factor and distances were small. Each of the three 1 : 100,000 maps containing parts of the northeastern area were used as geographical strata within which environmental combinations based on mean annual temperature, mean annual rainfall, and lithology were mapped with a GIS. Sites were selected for second stage sampling by topographic position, as previously described (Austin & Heyligers, 1989).
3. Modeling. The data finally consisted of 9537 plots. After setting an acceptance criterion of at least 50 presence observations in order to include a species in the GAM modeling, 88 tree species were modeled. To reduce the time taken to model the species, the same model was fitted to all species. The eleven predictors used for *E. fastigata* (see above) were used. Use of such a generic model ignoring significance levels will result in overspecification. The degree to which this reduces the accuracy of the models is the subject of current research.
4. GIS. A GIS with a 1-ha resolution was available containing all the necessary predictor variables, so predictions for each of the 88 species was possible for each of the 2.7 million cells in the GIS in the 27,000 km² zone.
5. Classification. To provide a community classification of the zone, the 2.7 million pixels, each characterized by the probability of occurrence of 88 tree species, were used in a numerical classification using ALOC and UPGMA procedures in the package PATN (Belbin, 1995). The available computer facilities and time imposed major limitations on the analysis of the large species by pixel matrix.

The final stage was a manual reorganization of the classification dendrogram to provide mappable units. Vegetation composition is strongly controlled by aspect in the area, and classification units were grouped into catenary sequences to give spatially coherent units for mapping. Two levels of vegetation classification were recognized: classes roughly corresponding to formations or alliances, and units approximating communities. Figure 9 gives examples of the class maps obtained. These maps at the finer scale of units, when combined with a land-cover map showing remaining forest areas, were used to decide that a further 100,000 ha of forest needed to be reserved in order to conserve an adequate representation of the pre-1750 forest communities. Pressey (in press) discusses the relationship between the scientific analysis and the political process for a similar exercise for northern NSW.

When information of this kind is available, then a further stage is reached where methods are needed to determine biodiversity priority areas. Explicit criteria are required, but the particular techniques are dependent on the available data. In general terms there are two classes of methods: (1) those that identify a set of areas in which all selected biodiversity attributes (e.g., species) are represented a specified number of times, e.g., once, twice, or three times; (2) those that maximize the amount of biodiversity represented by a given number of areas (Margules & Redhead, 1995). In the first the level of representation is specified arbitrarily, while in the second it is the number of areas that is fixed. This has become a major area of research and innovation where numerous constraints and trade-offs have been incorporated into the computer algorithms. In Australia, Margules and Nicholls (1987) pioneered an effective computer algorithm for the attribute-representation approach. Subsequently these authors with Pressey explored a number of the options with this approach (Nicholls & Margules, 1993; Pressey & Nicholls, 1989a, b; Pressey, 1994). Faith (1994) has developed a number of approaches to the second class of methods, using measures of dissimilarity and ordination techniques (Faith & Walker, 1994, 1997; Faith & Nicholls, 1997). Two features of the work by these authors are worthy of comment. First, the recognition that species richness per se is not a good criterion for conserving representative biodiversity; it is easily shown with simple examples that selecting the richest site of three may result in conserving fewer species than selecting the two sites each with fewer species. Complementarity of site composition is more important than maximal richness of individual sites. Second, the recognition that sophisticated algorithms are only valuable if they can be used with the limited and arbitrary data sets currently available and enhance those data rather than hide their inadequacies. The BioRap manuals (Margules & Redhead, 1995; Boston, 1997; Hutchinson et al., 1997; Faith & Nicholls, 1997; Noble, 1997) provide case studies of the use of alternative methods with various types of data.

Discussion

Herbarium records, while a primary source of data, have their limitations for analysis of species distributions (Hall, 1994; Austin et al., 1994a; Soberón et al., 1996). Data quality is a key issue, and computer routines for examining records of a species' distribution are an important first step (Chapman & Busby, 1994). New approaches such as BIOCLIM (Nix, 1986; Busby, 1991) and HABITAT (Walker & Cocks, 1991) are examples of heuristic methods designed to overcome the limitations of presence data. One difficulty is that survey data will age taxonomically. Without voucher specimens, it will not be possible to update survey records to take account of taxonomic revisions. However, managing biological diversity will require better data than herbarium presence records provide. Haila and Margules (1996) argued strongly that a necessary component of any practicable strategy for preserving the biological diversity of the earth is systematic field survey. They noted, however, that modern theoretical ecologists regard surveys as tedious, mundane activities; yet such data are essential to testing theory. Any survey has implicit in its design a set of ecological assumptions and a set of statistical assumptions; if these are not recognized and progressively improved upon, then maximum use will not be made of our limited survey resources. This paper has attempted to present some of the Australian experience in this area, but rapid changes are occurring as a result of society's demands that decisions be made on the inadequate database that currently exists. Poor survey design and predictive modeling techniques are adding to the difficulties. A major reason for this is that much of the work is appearing in the "gray" literature, and is inaccessible to many conservation scientists who might otherwise use the improved methods and techniques if they were aware of them. This review suffers from this problem in that much of the Australian work, good and bad, has yet to be published in the international literature and only exists in internal reports or reports published with small numbers of copies. Electronic publication may solve this problem of access to the literature.

Computer technology in various forms, remote-sensing, GIS, and statistical software are being used to create new tools for the study of biodiversity. What is more important is that we are finding new ways of thinking about the problems of studying biodiversity, whether it is how to design surveys or to develop new theories integrating ecology and evolution to better conserve our flora. Each stage in the study of biodiversity is now the subject of intense investigation in terms of basic research, conservation application, and cost-effectiveness (Margules & Austin, 1991). In addition, the results of such biodiversity studies are being incorporated into computer packages designed to facilitate community decision-making in regional land-use plans (Cocks et al., 1995) and are being actively used in conservation planning (Pressey, in press). A period of evaluation is now needed to determine which of these methods or tools are the best.

At the present time our immediate pragmatic concern is to make the best possible use of the biodiversity data we currently have to make sensible conservation decisions. Margules and his colleagues, in putting together the BioRap manuals and software for rapid assessment of biodiversity priority areas for the World Bank (with funding from AustAid), have shown how to make use of available data. The opportunity to constantly reiterate the processes is one of the strongest arguments for having computer-based tools for all aspects of biodiversity study: they can be repeated when necessary.

Literature Cited

Aleksander, I. & H. Morton. 1990. An Introduction to Neural Computing. Chapman & Hall, London.

Austin, M. P. 1992. Modeling the environmental niche of plants: Implications for plant community response to elevated CO_2 levels. Austral. J. Bot. 40: 615–630.

——— & J. J. Basinski. 1979. Bio-physical survey techniques. Pp. 24–42 *in* M. P. Austin & K. D. Cocks (general editors), Land Use on the South Coast of New South Wales, Vol. 1. CSIRO, Melbourne.

——— & E. M. Cawsey. 1991. Sampling strategies costed by simulation. Pp. 167–175 *in* C. R. Margules & M. P. Austin (editors), Nature Conservation: Cost Effective Biological Surveys and Data Analysis. CSIRO, Melbourne.

——— & M. J. Gaywood. 1994. Current problems of environmental gradients and species response curves in relation to continuum theory. J. Veg. Sci. 5: 473–482.

——— & P. C. Heyligers. 1989. Vegetation survey design for conservation: Gradsect sampling of forests in North-eastern New South Wales. Biol. Conservation 50: 13–32.

——— & ———. 1991. Vegetation survey design, a new approach: Gradsect sampling. Pp. 31–36 *in* C. R. Margules & M. P. Austin (editors), Nature Conservation: Cost Effective Biological Surveys and Data Analysis. CSIRO, Melbourne.

——— & J. A. Meyers. 1995. Modeling of Landscape Patterns and Processes Using Biological Data. Subproject 4: Real Data Dase Study. Consultancy Report for ERIN, CSIRO Division of Wildlife & Ecology, Canberra.

——— & ———. 1996. Current approaches to modeling the environmental niche of eucalypts: Implications for management of forest biodiversity. Forest Ecol. Managem. 85: 95–106.

——— & A. O. Nicholls. 1997. To fix or not to fix the species limits, that is the ecological question: Response to Jari Oksanen. J. Veg. Sci. 8: 743–748.

——— & T. M. Smith. 1989. A new model for the continuum concept. Vegetatio 83: 35–47.

———, A. O. Nicholls & C. R. Margules. 1990. Measurement of the realized qualitative niche: Environmental niches of five *Eucalyptus* species. Ecol. Monogr. 60: 161–177.

———, J. A. Meyers & M. D. Doherty. 1994a. Modeling of Landscape Patterns and Processes using Biological Data. Sub-project 2: Predictive Models for Landscape Patterns and Processes. Consultancy Report for ERIN, CSIRO Division of Wildlife and Ecology, Canberra.

———, J. G. Pausas & A. O. Nicholls. 1996. Patterns of tree species richness in relation to environment in south–eastern New South Wales, Australia. Austral. J. Ecol. 21: 154–164.

———, A. O. Nicholls, M. D. Doherty & J. A. Meyers. 1994b. Determining species response functions to an environmental gradient by means of a ß-function. J. Veg. Sci. 5: 215–228.

———, J. A. Meyers, D. L. Belbin & M. D. Doherty. 1995. Modeling of Landscape Patterns and Processes using Biological Data. Subproject 5: Simulated Data Case Study. Consultancy Report for ERIN, CSIRO Division of Wildlife & Ecology, Canberra.

Belbin, D. L. 1995. PATN Users Guide and Technical References. CSIRO Division of Wildlife and Ecology, Canberra.

Boston, A. N. 1997. BioRap. Rapid Assessment of Biodiversity. Vol. 1. The BioRap Biological Database. Australian BioRap Consortium, CSIRO, Canberra.

Breiman, L., J. H. Friedman, R. A. Olshen & C. J. Stone. 1984. Classification and Regression Trees. Wadsworth International Group, Belmont, California.

Burbidge, A. A. 1991. Cost constraints on surveys for nature conservation. Pp. 3–6 *in* C. R. Margules & M. P. Austin (editors), Nature Conservation: Cost Effective Biological Surveys and Data Analysis. CSIRO, Melbourne.

Busby, J. R. 1986. A biogeoclimatic analysis of *Nothofagus cunninghamii* (Hook.) Oerst. in southeastern Australia. Austral. J. Ecol. 11: 1–7.

———. 1988. Potential impacts of climate change on Australia's flora and fauna. Pp. 387–398 *in* G. I. Pearman (editor), Greenhouse: Planning for Climate Change. CSIRO, Melbourne.

———. 1991. BIOCLIM—A bioclimate analysis and prediction system. Pp. 64–68 *in* C. R. Margules & M. P. Austin (editors), Nature Conservation: Cost Effective Biological Surveys and Data Analysis. CSIRO, Melbourne.

Chapman, A. D. & J. R. Busby. 1994. Linking plant species information to continental biodiversity inventory, climate modeling and environmental monitoring. Pp. 179–195 *in* R. I. Miller (editor), Mapping the Diversity of Nature. Chapman & Hall, London.

Cocks, K. D., J. R. Ive & J. L. Clark. 1995. Forest Issues. Processes and Tools for Inventory, Evaluation, Mediation and Allocation. Report of a Case-study of the Bateman's Bay Area, New South Wales, Australia. CSIRO Division of Wildlife and Ecology Project Report, Canberra.

Currie, D. J. 1991. Energy and large-scale patterns of animal- and plant-species richness. Amer. Naturalist 137: 27–49.

Faith, D. P. 1994. Phylogenetic pattern and the quantification of organismal biodiversity. Philos. Trans., Ser. B. 345: 45–58.

——— & P. A. Walker. 1994. DIVERSITY: A Software Package for Sampling Phylogenetic and Environmental Diversity. Reference Manual and User's Guide. v.2.1. CSIRO Division of Wildlife and Ecology, Canberra.

——— & ———. 1997. Environmental diversity: On the best-possible use of surrogate data for assessing the relative biodiversity of sets of areas. Biodiversity & Conservation 5: 399–415.

——— & A. O. Nicholls. 1997. BioRap Rapid Assessment of Biodiversity, Vol. 3. Tools for Assessing Biodiversity Priority Areas. Australian BioRap Consortium. CSIRO, Canberra.

Fitzgerald, R. W. & B. G. Lees. 1992. The Application of Neural Networks to the Floristic Classification of Remote Sensing and GIS Data in Complex Terrain. Proceedings of the XVII Congress of the International Society for Photogrammetry and Remote Sensing, Washington, D.C.

Franklin, J. 1995. Predictive vegetation mapping: Geographic modeling of biospatial patterns in relation to environmental gradients. Progr. Phys. Geogr. 19: 474–499.

Gillison, A. N. & K. R. W. Brewer. 1985. The use of gradient directed transects or gradsects in natural resource survey. J. Environm. Managem. 20: 103–127.

Green, M. J. B. & E. R. N. Gunawardena. 1993. Conservation Evaluation of Some Natural Forests in Sri Lanka. Forestry Department, Sri Lanka, in assocation with UNDP, FAO, and IUCN.

Greig-Smith, P. 1983. Quantitative Plant Ecology, 3rd. ed. Blackwell Scientific Publications, Oxford.

Hall, J. B. 1994. Mapping for monographs: Baselines for resource development. Pp. 21–35 *in* R. I. Miller (editor), Mapping the Diversity of Nature. Chapman & Hall, London.

Haila, Y. & C. R. Margules. 1996. Survey research in conservation biology. Ecography 19: 323–331.

Hastie, T. & R. Tibshirani. 1990. Generalised Additive Models. Chapman & Hall, London.

Holland, J. H. 1992. Genetic algorithms. Sci. Amer. 267(1): 44–50.

Hutchinson, M. F. 1984. A Summary of Some Surface Fitting and Contouring Programs for Noisy Data. Consulting Report No. ACT 84/6, CSIRO Division of Mathematics and Statistics and Division of Water and Land Resources. CSIRO, Canberra.

——— & R. J. Bischof. 1983. A new method for estimating the spatial distribution of mean seasonal and annual rainfall applied to the Hunter Valley, New South Wales. Austral. Meteorol. Mag. 31: 179–184.

———, L. Belbin, A. O. Nicholls, H. A. Nix, J. P. McMahon & K. D. Ord. 1997. BioRap Rapid Assessment of Biodiversity, Vol. 2. Spatial Modeling Tools. Australian BioRap Consortium, CSIRO, Canberra.

Jongman, R. G. H., C. J. F. ter Braak & O. F. van Tongeren. 1987. Data Analysis in Community and Landscape Ecology. Pudoc, Wageningen.

Kent, M. & P. Coker. 1992. Vegetation Description and Analysis: A Practical Approach. Belhaven Press, London.

Leathwick, J. R. 1995. Climatic relationships of some New Zealand forest tree species. J. Veg. Sci. 6: 237–248.

——— & N. D. Mitchell. 1992. Forest pattern, climate and vulcanism in central North Island, New Zealand. J. Veg. Sci. 3: 603–616.

Lees, B. G. 1994. Decision Trees, Artificial Neural Networks and Genetic Algorithms for Classification of Remotely Sensed and Ancillary Data. 7th Australian Remote Sensing Conference Proceedings, Melbourne, Australia, March 1994.

——— & K. Ritman. 1991. Decision-tree and rule-induction approach to integration of remotely sensed and GIS data in mapping vegetation in disturbed or hilly environments. Environm. Managem. 15: 823–831.

Longmore, R. (editor). 1986. Atlas of Elapid Snakes of Australia. Australian Flora and Fauna Series, Number 7. Bureau of Flora and Fauna, Australian Government Publishing Service, Canberra.

McCullagh, P. & J. A. Nelder. 1989. Generalized Linear Models, 2nd ed. Chapman & Hall, London.

McKenzie, N. L., A. C. Robinson & L. Belbin. 1991. Biogeographic survey of the Nullarbor District, Australia. Pp. 54–63 *in* C. R. Margules & M. P. Austin (editors), Nature Conservation: Cost Effective Biological Surveys and Data Analysis. CSIRO, Melbourne.

Margules, C. R. & M. P. Austin (editors). 1991. Nature Conservation: Cost Effective Biological Surveys and Data Analysis. CSIRO, Melbourne, Australia.

——— & ———. 1994. Biological models for monitoring species decline: The construction and use of data bases. Philos.Trans., Ser. B 344: 69–75.

——— & A. O. Nicholls. 1987. Assessing the conservation value of remnant habitat 'islands': Mallee patches on the western Erye Peninsula, South Australia. Pp. 89–102 *in* D. A. Saunders, G. W. Arnold, A. A. Burbidge & A. J. M. Hopkins (editors), Nature Conservation: The Role of Remnant Vegetation. Surrey Beatty and Sons, Sydney.

——— & T. R. Redhead. 1995. BioRap. Guidelines for Using the BioRap Methodology and Tools. CSIRO Dickson, Canberra.

Neave, H. M., T. W. Norton, R. B. Cunningham & H. A. Nix. 1996. Biological inventory for conservation evaluation: I. Design of a field survey for diurnal, terrestrial birds in southern Australia. Forest. Ecol. Managem. 85: 107–122.

Nicholls, A. O. 1989. How to make biological surveys go further with Generalised Linear Models. Biol. Conservation. 50: 51–75.

———. 1991. Examples of the use of generalized linear models in analysis of survey data for conservation evaluation. Pp.191–201 *in* C. R. Margules & M. P. Austin (editors), Nature Conservation: Cost Effective Biological Surveys and Data Analysis. CSIRO, Melbourne.

——— & C. R. Margules. 1993. An upgraded reserve selection algorithm. Biol. Conservation 64: 165–169.

Nix, H. 1986. A biogeographic analysis of Australian elapid snakes. Pp. 4–15 *in* R. Longmore (editor), Atlas of Elapid Snakes of Australia. Australian Flora and Fauna Series, Number 7. Bureau of Flora and Fauna, Australian Government Publishing Service, Canberra.

——— & M. A. Switzer (editors). 1991. Rainforest Animals: Atlas of Vertebrates Endemic to Australia's Wet Tropics. Kowari No. 1. Australian National Parks and Wildlife Service, Canberra.

Noble, I. R. 1989. Ecological traits of *Eucalyptus* L'Herit. subgenera *Monocalyptus* and *Symphyomyrtus*. Austral. J. Bot. 37: 207–224.

Noble, S. J. 1997. BioRap Rapid Assessment of Biodiversity, Vol. 4. Tools for Storing and Mapping Spatial Data. Australian BioRap Consortium, CSIRO, Canberra.

Norton, T. W. & N. D. Mitchell. 1993. Application of generalised additive models to wildlife modeling and population viability analysis. *In:* Proceedings of the 7th Annual Symposium on Geographic Information Systems in Forestry, Environments and Natural Resources. Vancouver, Canada, February 1993. Ministry of Supply and Services, Vancouver, Canada.