PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Biological Models for Monitoring Species Decline: The Construction and Use of Data Bases [and Discussion]

C. R. Margules, M. P. Austin, D. Mollison and F. Smith

Phil. Trans. R. Soc. Lond. B 1994 **344**, 69-75
doi: 10.1098/rstb.1994.0053

Biological models for monitoring species decline: the construction and use of data bases

C. R. MARGULES AND M. P. AUSTIN

CSIRO Division of Wildlife and Ecology, P.O. Box 84, Lyneham, ACT 2602, Australia

SUMMARY

We describe procedures for designing ecological surveys and analysing survey data which more accurately estimate the spatial distribution patterns of species than the opportunistic or *ad hoc* biological field collections generally available. Surveys should be based on an environmental stratification, and the data collected should be sufficient to enable statistical estimates of wider spatial distribution patterns to be made from the point sample sites of the survey itself. Without such ecological data bases of high quality, accurate estimates of extinction rates will remain problematical, that is, confined to single populations and geographically restricted.

1. INTRODUCTION

A necessary prerequisite for estimating extinction rates is a data base that adequately represents the biota and is comprehensive in its geographical coverage. This is a fundamental and obvious requirement, so obvious, in fact, that it seems often to be taken for granted that such data exist. In reality, only a few small parts of the world with a long history of biological field surveys, such as the British Isles, can lay claim to adequate and comprehensive data bases, and even these can only be considered adequate for some components of the biota at some scales. Usually, field records are collected in a haphazard or opportunistic manner, the species recorded are the ones of interest to the collector, and the places from which they are recorded are the places those species might be expected to be found, or are conveniently accessible. Accordingly, extensive and often detailed collections of field records in museums, herbaria and various natural resource management agencies throughout the world are flawed, because they are incomplete and often biased, both in geographical coverage and in the sense that they are records of subsets of taxa.

The purpose of this paper is to describe some recent advances in the design of biological surveys and the analysis of survey data to estimate more accurately the spatial distribution patterns of species, and thus improve the baseline against which extinction rates can be estimated.

To establish a sound ecological data base requires (i) a conceptual framework based on ecological theory; (ii) field survey design principles based explicitly on the conceptual framework for locating field sample sites; (iii) a rationale for determining which measurements should be made at the chosen field sample sites in addition to records of the target species; and (iv) appropriate statistical methods for analysing survey data and predicting wider distribution and abundance patterns from the point records that the field sample sites represent.

Each of these requirements is considered in turn below.

2. THE CONCEPTUAL FRAMEWORK

Figure 1 is a map of koala records from part of the mid-north coast of New South Wales, Australia. Notice that these koala records closely map the road network. They are *ad hoc* and opportunistic. It is not possible, from this map, to define the range of this species. The sites with records are almost certainly a subset of the sites where this species actually occurs, and there are few, if any, records of where it was looked for but not found, that is, sites with recorded absences. The koala is probably Australia's most charismatic animal, but it is still not possible to define its range even in an area close to major population centres.

This is only one example of a pervasive problem with most existing data bases. Field records of trees in the Amazon, for instance, map the river network (G. Prance, personal communication). Such data bases contain recorded locations of some species but not all (or most) actual locations and no records of absences. They are not even representative of, let alone sample adequately, real distribution patterns.

Plant ecologists have adopted the concept of the individualistic continuum to explain observed patterns of variation in vegetation (Gleason 1926). The continuum concept holds that each species has a unique distribution, determined by its genetic make-up and physiological requirements, which is constrained by ecological interactions with other species. It is closely related to the niche concept used by animal ecologists (Whittaker *et al.*, 1973; Austin

Phil. Trans. R. Soc. Lond. B (1994) **344**, 69–75
Printed in Great Britain

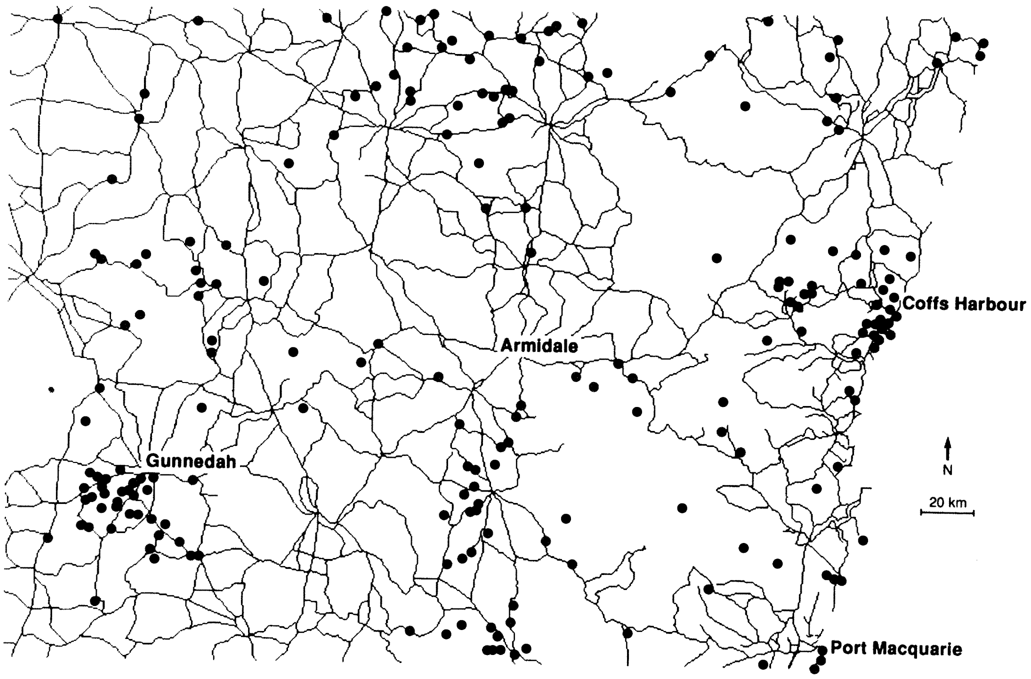

Figure 1. Koala records (courtesy of New South Wales National Parks & Wildlife Service) and the road network on part of the New South Wales north coast.

1985), and similar continuum patterns apply to animals (Rotenberry & Wiens 1980).

This is an appropriate conceptual framework for ecological survey design because it links species distribution patterns with variation in the environment. Whittaker (e.g. 1956, 1960) and Perring (e.g. 1958, 1959, 1960) provide a rationale for the types and scales of environmental gradients that could be incorporated in survey designs. More recently, Nix (1982) argued that, for the purpose of estimating distribution patterns of plants and animals, complete niche specification is not necessary, and that five environmental régimes, namely solar radiation, temperature, moisture, mineral nutrients and other components of the biota, are sufficient.

The goal of an ecological survey is to detect accurately species distribution patterns in both environmental and geographic space. The lesson for survey design is that the best estimates of which species occur in a region, and the patterns of abundance and range they exhibit, require the region to be stratified using major resource gradients or environmental variables, such as temperature, moisture and substrate, and to ensure that the range of combinations of these variables is sampled. Although the major determinants of patterns of occurrence are environmental, distributions are also conditional on co-occurring species. Thus it is necessary to record a range of species from survey sites to examine processes influencing persistence or extinction. Biological collections alone are insufficient for this purpose. Ecological data bases are needed.

3. FIELD SURVEY DESIGN

Survey data should be accurate and reliable, and fairly represent the true distribution and abundance patterns of the species recorded. Yet the design of surveys is a neglected topic (Austin & Heyligers 1991). Surveys themselves can be tedious, time consuming and labour intensive. Biological survey is not seen as a scientific endeavour and is therefore ignored by textbooks. There now exists a very sophisticated and still rapidly evolving technology for displaying and manipulating data in computers, but the methodology for acquiring those data in the first place remains primitive. The design of a survey has such profound implications for the subsequent use of the data that rigorous design rules should be formulated, explicated and applied.

Environmental stratification is an appropriate conceptual framework, but devising efficient and effective surveys, even within this framework, is an immensely practical problem. Gillison & Brewer (1985) proposed the use of gradient-directed transects, or gradsects, as a practical tool for designing surveys efficiently. The idea is to identify a set of transects which intercept the major environmental strata. If these transects (gradsects) are aligned along gradients of steep environmental change then the greatest amount of environmental variation can be intercepted in the shortest distance. Austin & Heyligers (1989) proposed refinements of this idea, including replication within transects and explicit rules for locating field sample sites, which incorporate another lower level of stratification.

(*a*) *An example: trees in coastal hardwood forest, New South Wales, Australia*

Austin & Heyligers (1989, 1991) conducted a survey of tree species in the *Eucalyptus* forests of coastal northern New South Wales. The area to be surveyed covered 20 000 km^2, the entire catchments of five major rivers. These rivers originate on tablelands and

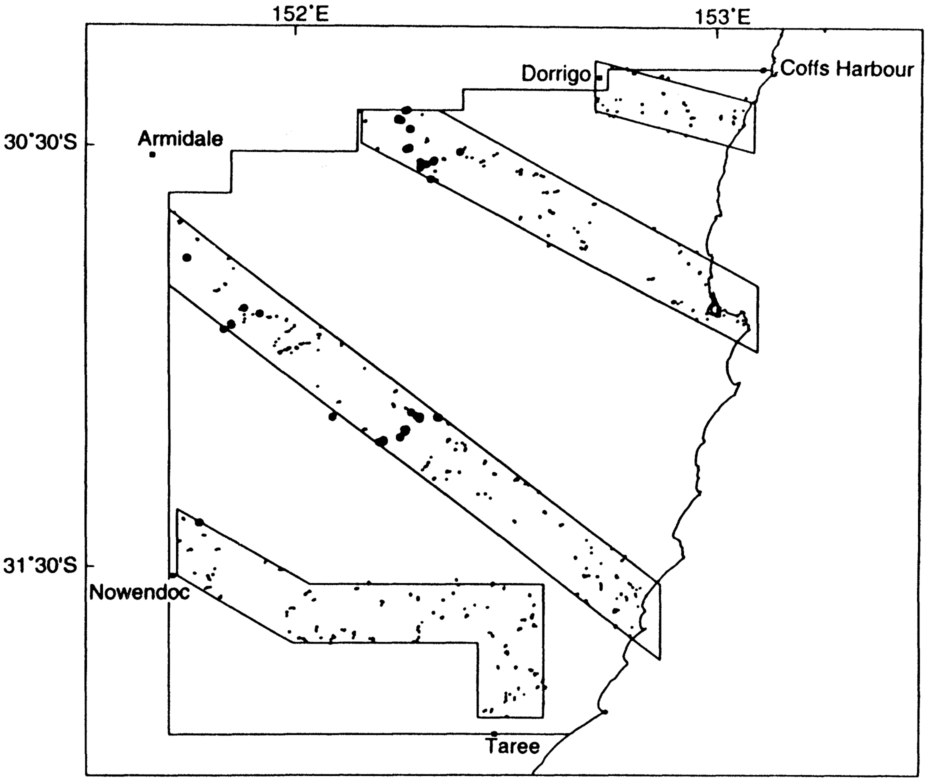

Figure 2. The north coast survey area showing the location of gradsects. The large dots represent survey sites where *Eucalyptus radiata* was found and the smaller dots all other survey sites (reproduced with permission of Elsevier Science Publishers from Nicholls (1989)).

flow east through deeply dissected mountains and hills onto a coastal plain and then into the Tasman Sea. Most of the area remains forested, although significant parts of the broader valleys and the coastal plain have been cleared for agriculture and settlement.

The Austin & Heyligers design protocol contained seven steps: (i) identify the major environmental variables influencing the distribution patterns of vegetation; (ii) recognize the subset of environmental variables best suited to determining the position and direction of gradsects so that they sample the range of combinations of environmental variables; (iii) choose the best available data for environmental stratification and the best available technology for implementing gradsect selection; (iv) stratify the environment within gradsects and break the gradsects up into segments to allow replicate sampling at different geographical locations; (v) decide whether or not another level of stratification is needed to take account of environmental variation at the local scale; (vi) decide on the effort that should be spent sampling the rarest environmental strata as opposed to increasing replication of the common strata; and (vii) be flexible; some sample sites selected in the laboratory will be useless because, for example, they have been cleared, or access is denied; new sites have to be chosen following established rules (suggested below).

(*b*) *Survey design*

Austin *et al.* (1984) had shown previously that rock type, rainfall and temperature have a strong influence on the distribution of tree species at this regional scale. As temperature correlates strongly with altitude, altitude was used because it could be easily determined in the field.

A regular 0.01 degree grid (approximately 1 km^2) was placed over the study area. Possible gradsect locations were evaluated by plotting all gridpoints against rainfall, altitude and rock type. These three major variables were divided into classes; 9 for rock type, 7 for altitude, and 8 for rainfall, producing $9 \times 7 \times 8 = 504$ possible combinations or environmental cells. In fact only 215 occur within the study area, mainly due to the localized distribution of rock types. Four was the maximum number of gradsects that could be sampled with available resources, and those four were positioned to encompass the maximum number of rainfall–altitude–rock-type combinations. Of the 43 combinations not covered, 18, nearly half, were represented in the study area by only one or two gridpoints. Figure 2 shows the selected gradsects, the distribution of sample points within them, and records of one species, which is used to exemplify analytical techniques below.

(*c*) *Sampling strategy*

Each gridpoint in each environmental cell should have an equal, or at least known, probability of being sampled. This is not possible in practice because access to many gridpoints would impose unacceptable costs (e.g. use of helicopters). Instead, Austin & Heyligers adopted a set of rules which ensured consistency of sampling and, being explicit, provided the opportunity for the degree of bias to be determined.

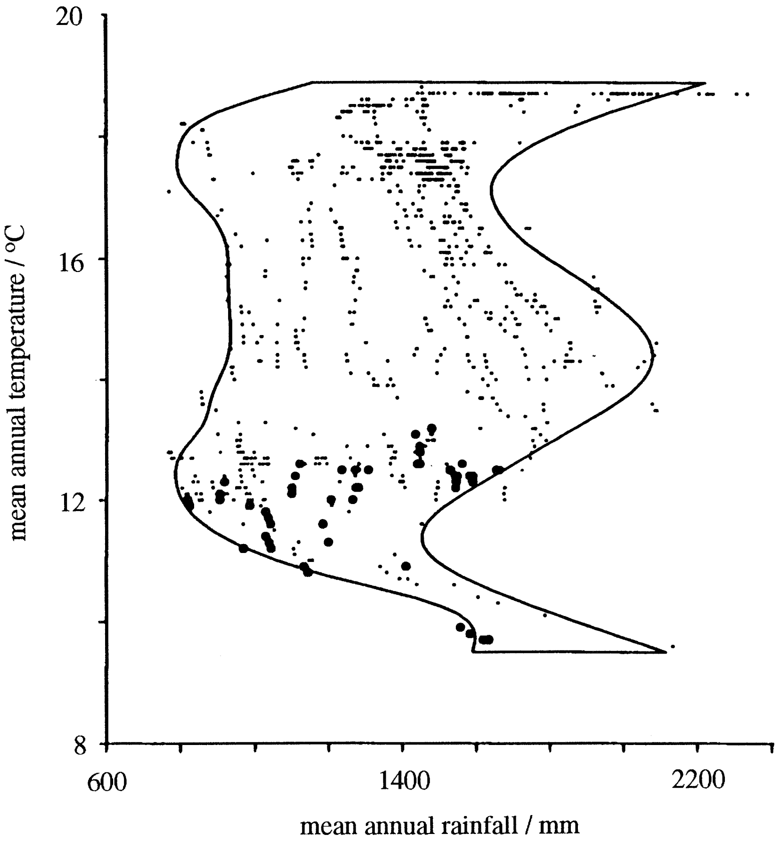

Figure 3. Survey sites plotted in the space formed by mean annual temperature and mean annual rainfall. The large dots represent sites where *Eucalyptus radiata* was recorded (reproduced with permission of Elsevier Science Publishers from Nicholls (1989)).

These included restricting sampling to within 0.5 km of access tracks, and proportional sampling depending on the number of gridpoints per environmental cell. Sample sites were selected randomly within each geographic segment. Each sample site consisted of five quadrats which together cover local environmental variation due to aspect and topographic position. Further details of quadrat location and field measurements can be found in Austin & Heyligers (1989, 1991).

4. FIELD MEASUREMENTS

The rationale for deciding on what to measure at each field sample site is determined by the requirement to predict wider geographic range and abundance patterns from the samples. Thus field measurements, aside from records of the target species, should be made of those variables most likely to correlate with species distribution patterns.

Austin *et al.* (1984, 1990), Margules & Nicholls (1987), Margules & Stein (1989) and Nicholls (1989, 1991) have successfully used environmental variables such as rainfall, temperature, lithological substrate and solar radiation to model the distribution patterns of plant species and plant communities. Braithwaite *et al.* (1984, 1989) and McKenzie *et al.* (1989) provide some empirical support for a similar relation between environmental variables and some animal distributions at a regional scale.

Ecological knowledge and experience have to be brought to bear. The most appropriate variables may differ in different ecosystems or biomes but the rationale is that they should be expected to be adequate predictors of wider distribution patterns. Austin & Heyligers (1989, 1991) recorded altitude, aspect, slope and topographic position which, when added to rainfall, temperature and rock type, the variables used in the survey design, were deemed to be appropriate and sufficient correlates of wider distribution patterns for trees in the forests of their study area, on the basis of previous experience and local knowledge.

5. ANALYTICAL METHODS

An appropriate analytical technique for survey data collected in an explicit systematic way is Generalized Linear Modelling (GLM) (McCullagh & Nelder 1983), which has traditional least-squares regression as a special case, is flexible in its assumptions, and allows the simultaneous use of continuous variables and factors. It is a tool that allows prediction from point samples via correlation with external variables to the wider geographic space. If the survey is complete, representing a true sample of the environmental space, the predictions will be interpolations with a high degree of confidence. If there are combinations of environmental variables not sampled in the survey then the predictions will be extrapolations beyond the domain of the data, and there will be less confidence in their accuracy.

Table 1. *Details of the environmental variables available for inclusion in a model developed to predict the probability of occurrence of* Eucalyptus radiata *by Nicholls (1989)*

variable	type	range or number of levels
altitude	continuous	0–1750 m
mean annual rainfall	continuous	800–2300 mm
mean annual temperature	continuous	9.0–19.9°C
lithology	categorical	9
topography	categorical	6
exposure	categorical	3

Nicholls (1989) modelled the distribution of one tree species from Austin & Heyligers' (1989, 1991) survey data, *Eucalyptus radiata*, which is reported here as an example. Other examples using different data sets can be found in Austin *et al.* (1984, 1990), Margules *et al.* (1987), Margules & Nicholls (1987) and Margules & Stein (1989). Leathwick & Mitchell (1992) provide a New Zealand example based on a conceptual framework which incorporates historical disturbance following volcanic activity.

Figure 2 shows the recorded geographical distribution of *E. radiata*, and figure 3 shows its environmental distribution in the space defined by rainfall and temperature. Table 1 lists the environmental variables available to Nicholls for use as predictors. Three of them are continuous and three are categorical.

Nicholls used the presence or absence of the species at a survey sample site as the dependent variable, and hence used logistic regression. He adopted a forward stepwise procedure, appropriate for an exploratory analysis of this kind. Complete details of the fitting procedure are supplied by Nicholls (1989).

The probability of *E. radiata* occurring at all combinations of rainfall, temperature and rock type found in the study area was calculated and the resulting surfaces contoured. A geographic map of the probability of occurrence of *E. radiata* was generated by calculating the probability of occurrence in each 1/100th degree grid cell using rainfall, temperature and the appropriate rock type. Figure 4 shows that map, which can be compared with figure 2.

6. DISCUSSION

Better data bases can be compiled with systematic stratified surveys based on a sound conceptual framework and the use of relevant analytical techniques. All of the methods are published and the software is widely available. The ideas behind them are essentially common sense. They were developed from the premise that complete inventories of regions or biomes are not a realistic option in the foreseeable future.

Collecting expeditions have not been conducted in the past with a view to mapping range and abundance

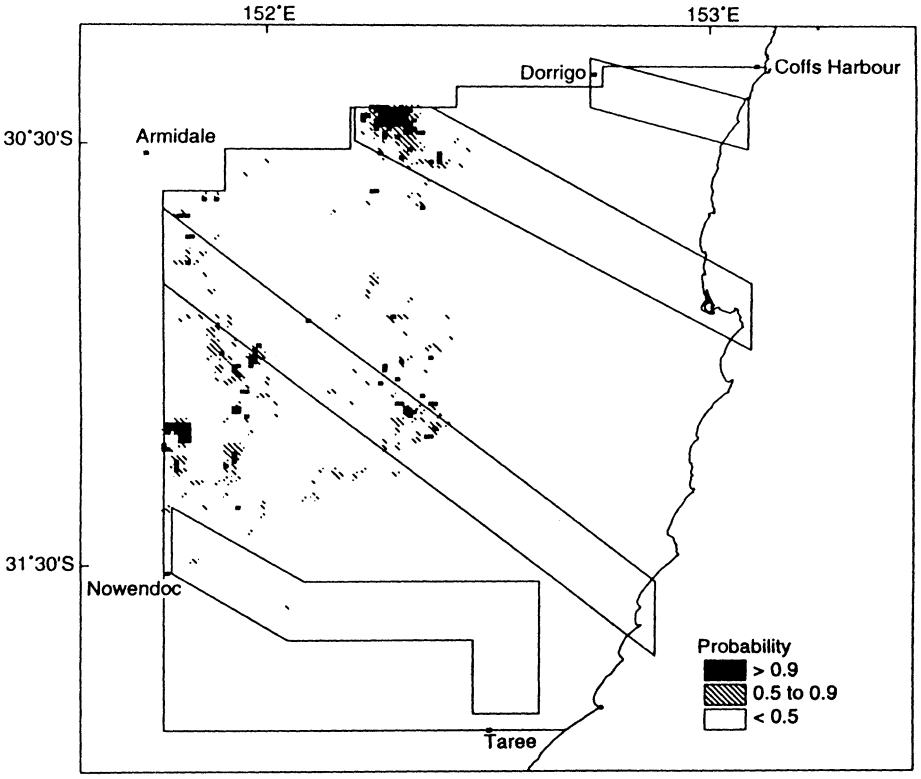

Figure 4. Predicted geographic distribution of *Eucalyptus radiata*, which can be compared with the recorded distribution within gradsects shown in figure 2 (reproduced with permission of Elsevier Science Publishers from Nicholls (1989)).

patterns of species, yet that is what collections in museums and herbaria are now being used for. Future collecting activities should incorporate environmental stratification in an explicit survey design. Green & Gunawardena (1993) have surveyed both the flora and fauna of Sri Lankan forests by using the gradsect design approach.

When the expense and labour invested in traditional collecting expeditions is considered, designed surveys will probably prove to be cheaper. Survey costs are almost never published (but see Margules & Austin 1991). Cost-effective surveys require both presence and absence data for predictive modelling (Margules & Austin 1991). However, there are empirical methods for estimating distribution patterns of species from presence-only data such as museum records. BIOCLIM is a well-known example (Nix 1986; Busby 1991) which compares the climatic profile of sites with known records of species with other sites to locate similar climatic profiles and therefore potential locations of those species. This is a rational way to use presence-only data, but the predictions do not have the same confidence that they would if the modelling technique was statistical. Museum records are also being used in an innovative way to help identify global priority areas for conservation action based on taxonomic diversity (Vane-Wright *et al.* 1991; Pressey *et al.* 1993).

Nevertheless, these are not reasons to continue with *ad hoc* or opportunistic collections of biological records. If future collections and other surveys were to be carefully planned based on environmental stratification and explicit design rules, the data could be used more effectively and with more confidence to estimate wider distribution patterns at a regional scale. One practical outcome would be to identify a subset of sites within a region that together sample, in a statistical sense, the biological diversity of that region as it is expressed in the data base (Margules & Nicholls 1994). This subset can be thought of as a nominal reserve network which samples regional diversity, and is therefore a suitable basis for developing regional conservation plans (Margules *et al.* 1994).

Although ecological surveys need a conceptual framework (see, for example, Austin & Smith 1989), they must also be informed by experience, intuition, common sense and local ecological knowledge. Ecological and evolutionary history play a role, in some cases a major role, in determining distribution patterns. Species with relict distributions or vicariant species occupying similar, but geographically isolated, environments will be detected with geographic replication but may not be modelled adequately. Plant species in the hyper-diverse areas of the Cape region, South Africa (Cowling 1992) and south-west Western Australia may be cases in point.

Nevertheless, a data base derived from ecological survey and predictive modelling using the idea of an environmental realized niche (Austin *et al.* 1990) as a conceptual framework is a first descriptive step towards understanding population dynamics in space and time, and, therefore, estimating more accurately probabilities of extinction.

Yrjö Haila and Albert van Jaarsveld commented critically on a draft of the manuscript, Paul Walker made the Koala data available, and Heather Lynch produced the maps. Jacqui Meyers drew the other figures.

REFERENCES

Austin, M.P. 1985 Continuum concept, ordination methods and niche theory. *A. Rev. Ecol. Syst.* **16**, 39–61.

Austin, M.P. 1987 Models for the analysis of species response to environmental gradients. *Vegetatio* **69**, 35–45.

Austin, M.P. & Heyligers, P.C. 1989 Vegetation survey design for conservation: gradsect sampling of forests in north-eastern New South Wales. *Biol. Conserv.* **50**, 13–32.

Austin, M.P. & Heyligers, P.C. 1991 New approach to vegetation survey design: gradsect sampling. In *Nature conservation: cost effective biological surveys and data analysis* (ed. C. R. Margules & M. P. Austin), pp. 31–36. Melbourne: CSIRO.

Austin, M.P. & Smith, T.M. 1989 A new model for the continuum concept. *Vegetatio* **83**, 35–47.

Austin, M.P., Cunningham, R.B. & Fleming, P.M. 1984 New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. *Vegetatio* **55**, 11–27.

Austin, M.P., Nicholls, A.O. & Margules, C.R. 1990 Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. *Ecol. Mongr.* **60**, 161–177.

Braithwaite, L.W., Austin, M.P., Clayton, M., Turner, J. & Nicholls, A.O. 1989 On predicting the presence of birds in *Eucalyptus* forest types. *Biol. Conserv.* **50**, 33–50.

Braithwaite, L.W., Turner, J. & Kelly, J. 1984 Studies of the arboreal marsupial fauna of eucalypt forests being harvested for woodpulp at Eden, New South Wales. III. Relationships between fauna densities, eucalypt occurrence and foliage nutrients and soil parent materials. *Aust. Wild. Res.* **11**, 41–48.

Busby, J.R. 1991 BIOCLIM – a bioclimatic analysis and prediction system. In *Nature conservation: cost effective biological surveys and data analysis*, (ed. C. R. Margules & M. P. Austin), pp. 64–68. Melbourne: CSIRO.

Caughley, G. 1994 Directions in conservation biology. *J. Anim. Ecol.* **63**.

Cowling, R.M. (ed.) 1992 *The ecology of fynbos*. Cape Town: Oxford University Press.

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

Gleason, H.A. 1926 The individualistic concept of the plant association. *Bull. Torrey bot. Club* **53**, 1–20.

Green, M.J.B. & Gunawardena, E.R.N. 1993 *Conservation evaluation of some natural forests in Sri Lanka.* Forestry Department, Sri Lanka, in association with UNDP, FAO and IUCN.

Leathwick, J.R. & Mitchell, N.D. 1992 Forest pattern, climate and vulcanism in central North Island, New Zealand. *J. Vegetation Sci.* **3**, 603–616.

Margules, C.R. & Austin, M.P. (ed.) 1991 *Nature conservation: cost effective biological surveys and data analysis.* Melbourne: CSIRO.

Margules, C.R., Cresswell, I.D. & Nicholls, A.O. 1994 A scientific basis for establishing networks of protected areas. In *Systematics and conservation evaluation* (ed. P. L. Forey, C. J. Humphries & R. I. Vane-Wright). Oxford University Press. (In the press.)

Margules, C.R. & Nicholls, A.O. 1987 Assessing the conservation value of remnant habitat 'islands': mallee patches on the western Eyre Peninsula, South Australia.

In *Nature conservation: the role of remnants of native vegetation* (ed. D. A. Saunders, G. W. Arnold, A. A. Burbidge & A. J. M. Hopkins), pp. 89–102. Chipping Norton, New South Wales: Surrey Beatty in association with CSIRO and CALM.

Margules, C.R. & Nicholls, A.O. 1994 Where should nature reserves be located? In *Conservation biology in Australia and Oceania* (ed. C. Moritz, J. Kikkawa & D. Doley). Chipping Norton, New South Wales: Surrey Beatty. (In the press.)

Margules, C.R. & Stein, J.L. 1989 Patterns in the distributions of species and the selection of nature reserves: an example from *Eucalyptus* forests in south-eastern New South Wales. *Biol. Conserv.* **50**, 219–238.

McCullagh, P. & Nelder, J.A. 1983 *Generalised Linear Models.* London: Chapman & Hall.

McKenzie, N.L., Belbin, L., Margules, C.R. & Keighery, G.J. 1989 Selecting representative reserve systems in remote areas: a case study in the Nullarbor region, Australia. *Biol. Conserv.* **50**, 239–261.

Nicholls, A.O. 1989 How to make biological surveys go further with generalised linear models. *Biol. Conserv.* **50**, 51–75.

Nicholls, A.O. 1991 Examples of the use of generalised linear models in analysis of survey data for conservation evaluation. In *Nature conservation: cost effective biological surveys and data analysis* (ed. C. R. Margules & M. P. Austin), pp. 54–63. Melbourne: CSIRO.

Nix, H.A. 1982 Environmental determinants of biogeography and evolution in Terra Australia. In *Evolution of the flora and fauna of arid Australia* (ed. W. R. Baker & P. J. M. Greenslade), pp. 47–66. Adelaide: Peacock Publishers.

Nix, H.A. 1986 A biogeographic analysis of Australian elapid snakes. In *Atlas of elapid snakes of Australia. Australian flora and fauna series, 7* (ed. R. Longmore), pp. 4–15. Canberra: Australian Government Publishing Service.

Perring, F. 1958 A theoretical approach to a study of chalk grassland. *J. Ecol.* **46**, 665–679.

Perring, F. 1959 Topographical gradients in chalk grassland. *J. Ecol.* **47**, 447–481.

Perring, F. 1960 Climatic gradients in chalk grassland. *J. Ecol.* **48**, 415–442.

Pressey, R.L., Humphries, C.J., Margules, C.R., Vane-Wright, R.I. & Williams, P.H. 1993 Beyond opportunism: key principles for systematic reserve selection. *Trends Ecol. Evol.* **8**, 124–128.

Whittaker, R.H. 1956 Vegetation of the Great Smoky Mountains. *Ecol. Monogr.* **26**, 1–80.

Whittaker, R.H. 1960 Vegetation of the Siskiyou Mountains, Oregon and California. *Ecol. Monogr.* **30**, 279–338.

Whittaker, R.H., Levin, S.A. & Root, R.B. 1973 Niche, habitat and ecotope. *Am. Nat.* **107**, 321–338.

Vane-Wright, R.I., Humphries, C.J. & Williams, P.H. 1991 What to protect? Systematics and the agony of choice. *Biol. Conserv.* **55**, 235–254.

Discussion

D. Mollison (*Heriot-Watt University, Edinburgh, U.K.*). Dr Margules and Dr Austin describe the contrasting community and continuum concepts for species niches. Have they analysed their data for individual species to see whether they provide evidence for or against the community concept? They seem to have an excellent data set for examining whether any pairs or groups of the tree species tend to occur together more or less often than would be expected purely from the environmental variables.

C. Margules. We have classified similar data into communities of co-occurring species which form spatial mapping units for the practical purposes of planning and management. Species with similar environmental requirements do tend to co-occur in space, but they are separated in environmental space, and as the environment changes, say, with latitude or altitude, some species drop out and others replace them. See Austin & Smith (1989) for a full discussion.

F. Smith (*Stanford University, U.S.A.*). The identification of regions in particular danger can be helped by overlaying multi-species data (e.g. mean range size, species richness) on topographic, climatic and particularly human data. In Australia, most mammal species are found between the east coast and the Great Dividing Range: these species have comparatively small ranges. Most human habitation is also in this region. There may be hotspots here which humans are unwittingly removing but it would be quite easy to identify them.

C. Margules. We have used similar models, including predictions of arboreal mammal distributions, to identify areas of high conflict between conservation interests and timber production interests in south-eastern Australia.