

# Spatial prediction of species distribution: an interface between ecological theory and statistical modelling

M.P. Austin

*CSIRO Sustainable Ecosystems, GPO Box 284, Canberra City, ACT 2601, Australia*

---

## Abstract

Neglect of ecological knowledge is a limiting factor in the use of statistical modelling to predict species distribution. Three components are needed for statistical modelling, an ecological model concerning the ecological theory to be used or tested, a data model concerning the collection and measurement of the data, and a statistical model concerning the statistical theory and methods used. This component framework is reviewed with emphasis on ecological theory. The expected shape of a species response curve to an environmental gradient is a central assumption on which agreement has yet to be reached. The nature of the environmental predictors whether indirect variables, e.g. latitude that have no physiological impact on plants, or direct variables, e.g. temperature also influence the type of response expected. Straight-line relationships between organisms and environment are often used uncritically. Many users of canonical correlation analysis use linear (straight-line) functions to relate ordination axes to variables such as slope and aspect though this is not a necessary part of the method. Some statisticians have used straight lines for species/environment relationships without testing, when evaluating new statistical procedures. Assumptions used in one component often conflict with those in another component. Statistical models can be used to explore ecological theory. Skewed species response curves predominate contrary to the symmetric unimodal curves assumed by some statistical methods. Improvements in statistical modelling can be achieved based on ecological concepts. Examples include incorporating interspecific competition from dominant species; more proximal predictors based on water balance models and spatial autocorrelation procedures to accommodate non-equilibrium vegetation.

© 2002 Elsevier Science B.V. All rights reserved.

*Keywords:* Vegetation models; Species models; Continuum; Niche; Generalised linear models; GLM; Generalised additive models; GAM; Canonical correspondence analysis; CCA; Spatial prediction; Regression; Ordination; Species response curves

---

## 1. Introduction

The spatial prediction of species distributions from survey data has recently been recognised as a significant component of conservation planning (Franklin, 1995; Austin, 1998, 2002; Guisan and

Zimmermann, 2000; Elith and Burgman 2002; Scott et al., 2002). A wide variety of statistical and machine-learning methods have been introduced, often in conjunction with geographic information systems (GIS) and remote-sensing (Fitzgerald and Lees, 1992; Aspinall and Veitch, 1993; Pereira and Itami, 1991; Franklin et al., 2000). Guisan and Zimmermann (2000) provide an extensive review of these developments, identifying

---

*E-mail address:* [mike.austin@csiro.au](mailto:mike.austin@csiro.au) (M.P. Austin).

many of the alternative statistical approaches that may be used. Various forms of regression analysis predominate in the literature. Generalised linear models (GLM, McCullagh and Nelder, 1989; Austin and Cunningham, 1981) and generalised additive models (GAM, Hastie and Tibshirani, 1990; Yee and Mitchell, 1991) with logistic regression using presence–absence survey data appear to be increasingly popular as the statistical model to be used (Franklin, 1995; Guisan and Zimmermann, 2000; Scott et al., 2002).

In the majority of cases, the purpose of the statistical modelling is the prediction of species distribution. The detection of functional relationships between species and environment and the testing of ecological theory tend to be secondary considerations (Guisan and Zimmermann, 2000). It will be argued in this paper that neglect of ecological knowledge is a limiting factor in the application of statistical modelling in ecology and conservation planning. At the interface between ecology and statistics, it is possible for statisticians to assume inadequate ecological models that may confound their evaluation of new statistical methods. On the other hand, ecologists may construct simpler statistical models than they believe are necessary because they are unaware of the power of modern statistical methods. A synthesis of current ecological theory and modern statistical models is badly needed. This paper reviews some of the current problems and tries to assemble some of the components necessary for a synthesis to be produced. The review is mainly concerned with plant community ecology, though similar arguments can be applied to other fields of ecology (Austin, 1999a,b).

## 2. A framework for statistical modelling in plant community ecology

Statistical models are based on correlation and often have as their purpose prediction. It is not possible to determine causation from correlation, but a description of functional relationships can be achieved (Sokal and Rohlf, 1981; Box, 1966). The reality of such relationships and the causal mechanisms responsible should then be pursued by

experiment, by theoretical analysis or by repeating the study at a different location. Another use of statistical models is calibration, where biotic data are used to estimate past environmental conditions, an approach of increasing importance in palaeoecology. Birks (1998) states “The most significant development in quantitative palaeolimnology has been the creation of many modern calibration data-sets of biotic assemblages and associated environmental data. Such calibration sets, when analysed by appropriate procedures, have the potential to transform fossil biostratigraphical data into quantitative estimates of the past environment.” These calibration methods (ter Braak and Verdonschot, 1995) are not considered further except where they relate to ecological assumptions.

There are three major components in any framework for statistical modelling in plant ecology. There needs to be an ecological model, a data model, and a statistical model. The ecological model consists of the ecological knowledge and theory to be used or tested in the study. The data model consists of the decisions made regarding how the data are collected and how the data will be measured or estimated. The statistical model involves the choice of statistical method, error function and significance tests. Each model interacts in both obvious and subtle ways with the other models to determine the success of any statistical modelling exercise.

The ecological model may contain assumptions to be incorporated into the analysis or hypotheses to be tested. It may, for example be assumed that vegetation and species composition are primarily determined by environment rather than by successional status or by time since last disturbance (Franklin et al., 2000). The functional link between plants and environment may be assumed to be all straight-line relationships (Austin, 1999a). A hypothesis might test whether species respond to environmental variables or gradients with a unimodal, symmetric, response function.

An important example of the data model is that used by classical phytosociologists (Braun-Blanquet, 1964; Westhoff and van der Maarel, 1978). Samples (plots) are selected subjectively by the ecologist without any statistical sampling design

considerations. Species abundances are estimated using a mixed cover/abundance scale with unequal class intervals, the statistical implications of which have not been considered until recently. Guisan and Harrell (2000, Guisan 2002) point out that this cover/abundance measurement is an ordinal variable that requires, therefore, special attention when selecting the statistical model. Franklin et al. (2000) consider a number of problems that can arise with the data model. These include problems of the appropriate scale of observations and locational accuracy when ecological field data are used with GIS and databases. The scale of resolution and types of errors associated with digital terrain models were also discussed. The authors included satellite imagery derived variables in their case studies. Such remotely-sensed variables have their own data models, which need to be evaluated for compatibility with ecological models. Franklin et al. (2000) provide a useful geographical perspective on issues associated with the choice of data model for vegetation prediction.

The term 'conceptual model' has been used to include both the ecological model and the data model. Separating theory from data availability and model applicability can ensure debates about theory are not confused with differences in data models. See the papers by Oksanen (1997), Austin and Nicholls (1997), Oksanen and Minchin (2002) for a recent example of a debate over theory and data models concerning the shape of species response curves.

Similar difficulties can arise with the statistical model regarding the error function assumed, choice of model whether generalised linear model (GLM), generalised additive model (GAM) or classification and regression tree (CART, Breiman et al., 1984) and methods of validation. Statistical decisions can change the ecological model. Choice of data transformation, e.g. log to stabilise the variance will change the functional relationship from an additive to a multiplicative form. Boyce et al. (2002) review statistical methods for evaluating resource selection functions (RSF) from the perspective of the ecological and data models used. RSF are defined as any function that is proportional to the probability of use of a resource or area by an organism. The approach was developed

for the selection of resources by animals but is very similar to estimating plant species response curves (Boyce et al., 2002). The ecological model and the data model for plants and animals differ. An important difference is the difficulty of estimating the non-use of a resource by an animal. Even estimating the absence from a habitat or site of a mobile or cryptic animal is a non-trivial problem. Consequently the statistical models may differ. However, Boyce et al.'s discussion of evaluation procedures has general applicability.

The focus here is on the many ecological assumptions and theories that can impact on spatial modelling of vegetation. Prediction can be achieved without the regression equation having any necessary ecological process basis, but the result is unlikely to be robust. The more that knowledge of ecological process can be incorporated, the more robust the prediction and the more explanatory power the equation is likely to have. However, statistical models currently used are static and contain no dynamic elements (Guisan and Zimmermann, 2000; Guisan and Theurillat, 2000). It is assumed that vegetation is in equilibrium with the environment, or at least a quasi-equilibrium where change is slow relative to the life span of the biota. This is a central ecological assumption of the use of these statistical models at the present time. O'Connor (2001) has also recently made this point. There are, therefore, limits on the degree of success that can be achieved by statistical models using environmental predictors. This will vary depending on the degree to which history and disturbance are important in the biome under study. Franklin and colleagues (Franklin et al., 2000; Meentemeyer et al., 2001) have recently shown the predictive importance of topographically dependent environmental gradients in the fire-prone California chaparral. Note the recent use of GAM by Fewster et al. (2000) to examine the population dynamics of birds and the use of temporal data by Boyce et al. (2002). Current techniques need not be limited to static, equilibrium situations but the nature of ecological, data and statistical models will need to be re-examined.

Where it is accepted that it is important to maximise our knowledge of the functional rela-

tionships between plants and environment, current ecological concepts need to be carefully considered. For example, consensus is emerging that the continuum concept is a more appropriate model of vegetation organisation than that of the plant community or association and is more suitable for spatial prediction (Austin and Smith, 1989; Franklin, 1995; Guisan and Zimmermann, 2000). Similarly, it is now well recognised that statistical models provide a description of the realised niche of a species but can say little about the fundamental niche (Austin and Smith, 1989; Austin, 1992; Malanson et al., 1992; Franklin, 1995; Guisan and Zimmermann, 2000). The statistical model may not represent the realised niche. The source-sink model (Pulliam, 1988) regards the geographical distribution of a species as composed of two areas. The source areas are where population growth is positive. Sink areas are those areas where population growth is below replacement and populations are maintained by dispersal from source areas. Hence statistical models may be describing an amalgam of realised niche and sink areas. Pulliam (1988) comments that ‘... in such cases, it can be said that the fundamental niche is smaller than the realised niche.’ Watkinson and Sutherland (1995) discuss some plant examples, *Cakite edentula* is given as an example where reproduction from populations at one position on an environmental gradient maintains sink populations at other positions on the gradient (Keddy, 1982; Watkinson, 1985). A more inclusive theoretical framework is needed which recognises both niche and source-sink ideas. Few statistical models of distribution have considered reproductive success though see Heegaard (2001) for an example of the occurrence of sink populations in an alpine moss.

There is much less consensus regarding other theoretical concepts in plant ecology than the continuum concept and their role in spatial modelling.

### 2.1. Biotic components

Assumptions about the shape of the response of species to an environmental variable (usually termed an environmental gradient) are central to

any predictive modelling effort. Most ecological textbooks (Begon et al., 1990; Giller, 1984; Krebs, 1994) present the response as a unimodal, symmetric bell-shaped curve. Austin (1999a) drew attention to the lack of evidence for this assumption. Niche theory is based on this assumption and additionally makes several others (Fig. 1):

- 1) Both the fundamental and realised niches of a species are bell-shaped symmetric curves.
- 2) Competition restricts niche breadth.
- 3) Maximum abundance occurs at the optimum for the fundamental niche.
- 4) Species maxima are equally spaced along the gradient.
- 5) Species maxima are of equal amplitude.
- 6) Collective properties of species, e.g. species richness, dominance or stand abundance show no patterns of response along the gradient.

Mueller-Dombois and Ellenberg (1974) Fig. 12.3, see also Austin (1999a) put forward a more general theory which relaxes several but not all of the above assumptions. Their theory assumes that a superior competitor can displace a species from the optimum of its fundamental niche. As a consequence, a species realised niche response may take a wide variety of shapes from skewed to bimodal (Fig. 1). There is no general agreement on species response shape. Many predictive studies

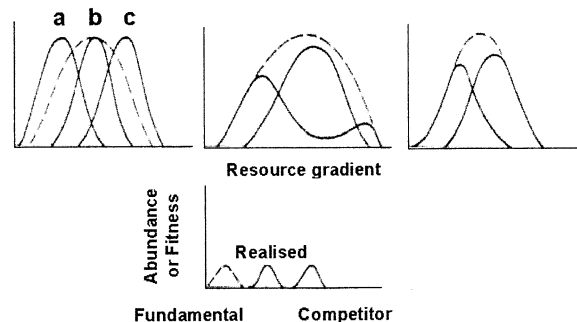


Fig. 1. Three alternative models of species response shape. The first diagram shows the niche theory model. The second and third diagrams show possibilities when a superior competitor displaces a species from its optimum while the fourth gives details of labels and curves. (Austin and Smith, 1989, with kind permission from Kluwer Academic Publishing.)

fail to address this issue when specifying the mathematical forms to be used in fitting their statistical models. Austin and Smith (1989) tabulated the assumptions adopted by many ecologists in their conceptual models of vegetation. Many failed to specify the shape of species responses, changes in total biomass along the gradient or species richness. Conceptual models of vegetation organisation are incomplete to judge from these tabulations. This may contribute to why many models of species distribution fail to test theoretical assumptions (Franklin 1995; Austin, 1999b; Guisan and Zimmermann, 2000; Scott et al., 2002).

While the continuum concept of vegetation may be accepted, little attention has been paid to what alternative patterns of species distributions might be compatible with the concept (Austin, 1985). Gleason's (1926) concept of species having individualistic distributions along the gradient though influenced by competition is only one possible realisation. Trees, shrubs and herbs may vary independently of each other but partition the environmental gradient according to conventional niche theory (Fig. 2). Other more complex possibilities can be imagined. If the response of shrubs and herbs is dependent on the composition of the canopy layer in forest this could have a profound

effect on the types of models needed to predict the distribution of shrubs and herbs in forests.

## 2.2. Environmental components

Theoretical discussion of environmental gradients does not usually distinguish between types of gradients (ter Braak, 1985, 1986; Austin and Smith, 1989; Guisan and Zimmermann, 2000). Two distinct classifications of environmental gradients have been used, idealised types and distal or proximal gradients.

Austin (1980) suggested there were three idealised kinds of gradients. These were direct, resource and indirect gradients. Indirect gradients are those where the variable has no physiological effect on plant growth or competition. Examples are altitude, latitude or longitude. Their correlation with species distribution is due to their location-dependent correlation with variables such as temperature or rainfall. Direct gradients are those that do have a direct physiological influence on growth but are not consumed by the plants. Examples are temperature and pH. Resource gradients are based on those variables that are consumed by plants such as light, water and nutrients. These are idealised and not exclusive categories. Water can be a resource gradient under conditions of low availability and an indirect gradient when it is sufficiently abundant to cause anaerobic conditions due to waterlogging.

Environmental predictors (gradients) may be either proximal or distal. Proximal and distal refer to the position of the predictor in the chain of processes that link the predictor to its impact on the plant. The most proximal gradient will be the causal variable determining the plant response. For example, available soluble soil phosphate concentration at the root hair would be a more proximal resource gradient than total soil phosphorus. Indirect gradients are clearly distal variables. Species distribution models will have only local value either for prediction or understanding when using distal variables. Models based on proximal resource and direct gradients will be the most robust and widely applicable. They will also be the least practical in terms of knowing what to measure or in terms of resources and time to

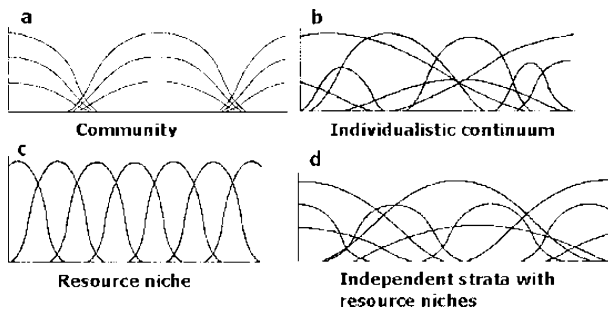


Fig. 2. Four alternative patterns of vegetation organisation along an environmental gradient. The a and b represent the classical community and continuum models. c Shows the niche theory model with species partitioning the resource while d shows a possible pattern where trees, shrubs and grasses partition the gradient independently. (Austin, 1985, with kind permission from Annual Reviews Inc.)

measure them. It will be very difficult to provide GIS coverage for proximal variables. Use of proximal variables for predictive mapping of species distribution is, therefore, probably impractical.

The expected shape of the species response will vary with the nature of the gradient. Response to an indirect gradient could take any shape. The shape would depend on the nature of the correlations between the indirect variable and the causal gradients (Guisan et al., 1999; Guisan and Zimmermann, 2000; Franklin et al., 2000; Meentemeyer et al., 2001). Guisan and Zimmermann (2000) point out that in many cases ‘indirect variables usually replace a combination of different resources and direct gradients in a simple way’. This is likely to lead to robust predictive models only where the causal gradients are linearly correlated and those gradients are themselves linearly correlated with the indirect gradient. Gradients of altitude at sub-alpine and alpine altitudes may show this characteristic. Topographic position is another indirect factor that may behave in this way, but once the lithology changes, the correlation structure between the variables will change. Austin and Smith (1989) discuss the expected response shape to direct and resource gradients. They proposed certain shapes but no definitive conclusions have been reached.

### 2.3. *Collective properties*

When plants grow together, the resulting community can be described by certain collective properties such as species richness, dominance and total biomass. The relationships between these properties have attracted considerable theoretical interest (Cornell and Lawton, 1992; Cornell and Karlson, 1997; Naeem et al., 1994, 1996; Tilman et al., 1996, 1997; Partel et al., 1996). Frequently, these studies use only simple regression models (Austin, 1999b). More complex statistical models have been used in exploratory analysis of species richness patterns in relation to environment (Currie, 1991; Margules et al., 1987; Pausas, 1994; Austin, et al., 1996). Investigations of global patterns of species richness tend to emphasise indirect gradients such as latitude (Ricklefs and

Schluter, 1993; Rosenzweig, 1995). A case can be made for more critical use of regression models when applied to collective properties of vegetation (Austin, 1999b; Pausas and Austin, 2001). Austin (1999b) comments that quadratic equations have been fitted where the authors’ theory indicates that a hyperbola should be used (Cornell and Karlson, 1997).

Guisan and Theurillat (2000) modelled species richness using a Poisson GLM. They then compared this pattern with that obtained by accumulating predictions from individual species. The results for current climatic conditions were similar. The patterns differed when different climate change scenarios were used with the models. Assumptions about the patterns of response of collective properties are also critical when simulated data are used to assess the performance of different types of statistical models (Minchin, 1987; Faith et al., 1987; Palmer, 1993; ter Braak et al., 1993; Austin et al., 1995). The relationship between patterns of individual species response curves and response curves of collective properties such as species richness remains to be elucidated.

### 2.4. *Ecological processes*

While the equilibrium assumption is essential for static modelling, other assumptions about ecological processes tend to remain unconsidered (Guisan and Zimmermann, 2000). Models of species distribution usually focus on environmental predictors. Processes such as dispersal, competition, succession, fire and grazing pressure need to be incorporated. Leathwick, in a series of studies on the distribution of New Zealand forest trees has incorporated a number of these features (Leathwick and Mitchell, 1992; Leathwick, 1995, 1998; Leathwick et al., 1996, 1998; Leathwick and Austin, 2001). Leathwick and Mitchell (1992) modelled the distribution of forest trees, in the central North Island of New Zealand using GLM. The region had been disturbed by the large Taupo Pumice eruption (130 AD). Depth of the pumice deposits was a significant predictor for several species. The positive correlation of rapidly dispersing conifers with the deeper deposits close to the eruption centre was interpreted as a result of a

dispersal process rather than an environmental control. All forests close to the eruption centre would have been destroyed. The slower dispersing *Nothofagus* species would normally dominate such sites but have yet to reach them. This is an example where spatial autocorrelation should be an intrinsic part of the model. Whenever a statistical model is established with spatial autocorrelation or autologistic components, one of two assumptions is being made (Wu and Huffer, 1997; Augustin et al., 1996; Gumpertz et al., 1997; Weir and Pettit, 2000). The model may have been mis-specified and an important environmental predictor showing spatial autocorrelation left out of the model. Alternatively, biological processes associated with species dispersal ability, historical disturbance or dispersal barriers such as mountain ranges may be responsible. Use of spatial autocorrelation in distribution models without considering the processes responsible can only lead to less than adequate models. Both Gumpertz et al. (1997), Weir and Pettit (2000) consider covariates but their impact on the methodological comparisons being made is not examined in detail. Augustin et al. (1996) include environmental covariates in their study of red deer distribution but admit that herd behaviour should perhaps have been included in the model. The interface between spatial autocorrelation methods and ecological processes needs greater attention.

Work using GAM by Leathwick (1995, 1998), Leathwick et al. (1996, 1998) on the phenomena of 'Beech Gaps' in New Zealand emphasised the need to combine ecological knowledge with statistical modelling. The gaps, which occur on both North and South Islands, do not correspond with any of the environmental factors currently known to influence tree species distribution in New Zealand (Leathwick, 1998). *Nothofagus* displacement is thought to have occurred due to major historical disturbance events, e.g. glacial advances, volcanic eruptions and plate tectonics (McGlone et al., 1996). Their slow reinvasion is probably due to their limited seed dispersal and obligate ectomycorrhizal association (Bayliss, 1980). In the 'Beech Gaps' where *Nothofagus* is absent, other tree

species behave differently in relation to environmental predictors. Leathwick and Austin (2001) introduced competition from the dominant tree species *Nothofagus* in modelling the spatial distribution of the density of these other forest species. Prediction was also improved by making the competitive effect of *Nothofagus* species a function of position on the environmental gradients (i.e. a term describing interaction between *Nothofagus* density and the environment value was used). The GAM results indicate substantial reductions in abundance of other tree species as *Nothofagus* abundance increases. The magnitude of the reduction varies with position on the dominant environmental gradient (Leathwick and Austin, 2001). Leathwick (2002) examines further the competitive influence of different *Nothofagus* species.

There are difficulties associated with introducing competitive terms into habitat modelling. It is difficult to distinguish whether absence and reduced abundance is due to competition or is due to an unidentified environmental variable. The *Nothofagus* example is exceptional. There is strong circumstantial evidence that the *Nothofagus* species are absent due to historical events. The probability of competitive interactions between species is much reduced with plot data unless the species are sufficiently large and abundant for them to interact as neighbours. When present the *Nothofagus* species are often the natural dominants and are the most abundant species on the plot. At a plot size of 0.4 ha interactions with other species are probable. The occurrence of density levels sufficient for interactions to occur will not be the case with less common species, reducing the value of this approach. Models based on presence-absence data are unlikely to detect such phenomena. The potential of simultaneous regression models for ecology (Brzeziecki, 1987) is likely to be restricted to very special circumstances for this reason. See discussion below of Austin (1971) for an alternative approach to incorporating competition. Biotic processes as distinct from environmental processes are rare in statistical models of species distributions and need greater consideration.

### 3. Use of ecological theory in plant community modelling

How is ecological theory being used in statistical models of plant species distribution? There are a number of problems at the interface with statistics that need to be considered more carefully by the analyst and perhaps applied more generally in ecology.

#### 3.1. Straight line models

Regression models fitting straight-line relationships between plant responses and environmental variables are frequent in the ecological literature. Statements that curvilinear relationships were tested are usually absent from such cases, even though a curvilinear relationship would make ecological sense.

An example typical of many in the literature might concern the number of flowering stems produced in a given year as a function of rainfall. This example may seem trivial to many ecologists. However, many multiple regressions consist solely of straight-line relationships without any concern for the ecological rationale of the relationship. Statistical significance is sufficient. A straight line was fitted to the data. The relationship was significant with  $r^2 = 0.30$  (Fig. 3). There is one noticeable and influential outlier. An alternative model, however, exists which is equally parsimonious, explains more of the variance and is ecologically more rational (Austin 1991). The

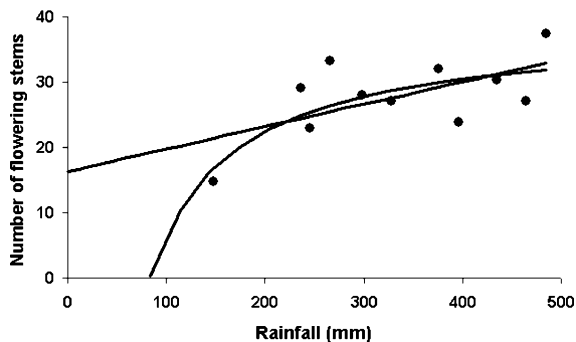


Fig. 3. An example of straight-line regression where an equally parsimonious and ecologically more rational model is possible. (From Austin, 1991, with permission of CSIRO Publishing.)

regression model  $y = a + b/x$  where  $x$  is annual rainfall achieves these outcomes ( $r^2 = 0.46$ ). The number of degrees of freedom is the same and the outlier is now incorporated in the model. An ecological interpretation of the model is that there is minimum rainfall below which no flowers are produced. Above this threshold, flowering increases with rainfall until a limit is reached above which rainfall has little or no effect.

Statisticians will regard the example as an example of poor data analysis. If examined, the pattern of deviations in the residuals would have exposed the limitations of the straight line. Yet statisticians when evaluating new approaches such as the Markov Chain Monte Carlo method on real data for species distributions do the same. Wu and Huffer (1997) model plant species distributions in Florida with straight-line relationships for climatic variables using logistic regression. They ascribe the pattern of residuals obtained to spatial autocorrelation in the species data. No mention is made of probable mis-specification of the model. Three methods of autologistic regression are then compared using the same data. Gumpertz et al. (1997) fit straight-line relationships for the effect of soil variables on spatial pattern of a disease. They mention the possibility of other relationships but do not pursue them. Weir and Pettit (2000) in methodological study of spatial autocorrelation using the common toad suggest the use of environmental covariates but only examine linear and quadratic spatial trend surface models. Latitude appears to be the indirect environmental gradient operating in this data set from Finland at the northern limits of the distribution of the common toad. Ecologically the model appears poorly specified. Temperature might be a more appropriate variable. What are the consequences of potentially mis-specifying the model on the evaluation of different statistical methods?

One of the most popular statistical methods in plant ecology at the present time is canonical correspondence analysis (CCA, ter Braak, 1986). The method combines a multivariate ordination of species occurrence data with a constrained regression maximising the correlation between the species ordination axes and selected environmental variables. The species are assumed to have unim-



odal responses to the underlying environmental gradients as specified by the ordination axes. Current practice is to estimate the underlying gradient as a linear (straight-line) combination of environmental variables. An arbitrary selection of 17 references from [Birks and Austin \(1992\)](#) bibliography of papers using CCA all used linear combinations. Many of the variables would be classified as distal indirect variables, e.g. latitude, altitude and aspect. A review of papers using CCA in volumes 10–12 of the *Journal of Vegetation Science* (1999–2001) gave 11 that did not mention the linear basis of the correlation. Twenty out of the 22 papers do not consider the possibility of curvilinear relationships or that distal indirect variables may have very different relationships with the ordination gradients than direct or resource variables. [Leps et al. \(2000\)](#) explicitly use interactions between variables in their constrained regression. However, they used principal components analysis (PCA) rather than CCA because the ordination axes were short as determined by detrended correspondence analysis (DCA). [Brunet et al. \(2000\)](#) is one of the few papers reviewed to provide graphical evidence of the relationship between the correspondence analysis (CA) axes and the environmental predictor variables. The main predictor was distance from the woodland edge; their figure one shows that for several sites the relationship between CA axis one and distance would be better described by a curvilinear function. [Ejrnæs and Bruun \(2000\)](#) using DCA found reciprocal and quadratic relationships for the DCA axes with pH and maximum summer irradiation, respectively.

There is no ecological reason to expect these variables to have linear relationships with the underlying direct and resource gradients that determine species patterns. There is no reason to expect that linear combinations of these variables will provide good estimates of the gradients either. The model of species response used in CCA is clear and theoretically acceptable to a degree, and the linkage with environment is explicit. In practice the measurement of relevant environmental variables is inadequate. The results will be subject to the same failure of interpretation as in the one variable straight-line example given above. This

need not be the case. [ter Braak \(1986, 1987\)](#) makes clear that various transformations and scalars can be used as the environmental predictors. Spline functions can be incorporated (W. Venables pers. com.). A failure in communication between statisticians and ecologists is occurring.

A detailed review of the impact of straight-line regressions on ecological interpretation is required.

### 3.2. *Assumed ecological theory*

Statistical regression models are not normally associated with an explicit ecological theory. One exception is CCA. [ter Braak \(1985, 1986\)](#) makes the ecological assumptions very clear:

- 1) the species tolerances (niche breadth) are equal;
- 2) the species maxima (amplitudes) are equal;
- 3) the species optima are homogeneously distributed along a gradient that is long compared with species niche breadth.

‘Homogeneously distributed’ refers to either equally spaced optima or optima distributed randomly from an uniform distribution. This implies a particular species-packing model ([ter Braak, 1986](#)). It is a one-dimensional model. Further axes are assumed to be independent of each other. The model is very similar to the standard niche theory discussed above. The assumptions define symmetric bell-shaped responses for the species. The influence of changes in species richness along gradients is not addressed; see [Pausas \(1994\)](#), [Austin et al. \(1995\)](#) for examples of well-defined patterns of species richness along environmental gradients. Changes in patterns of dominance, another collective property of vegetation, with associated changes in species amplitudes along gradients are not considered. The packing model must be considered incomplete ecologically in the sense of [Austin and Smith \(1989\)](#).

CCA is considered robust to changes in the assumptions ([Palmer, 1993](#); [ter Braak, 1987](#); [ter Braak et al., 1993](#); [ter Braak and Juggins, 1993](#)). [Palmer \(1993\)](#) showed that CCA recovers the distribution of vegetation plots in relation to the

true underlying gradients. To do this, the ordination axes were regressed on predictors consisting of the true gradients plus random variables. The outcome is not surprising; random variables are unlikely to be a match for the original generating variables. The success of CCA was not examined when indirect predictors with complex functional relationships with causal proximal variables are included. Ter Braak and colleagues (ter Braak et al., 1993; ter Braak and Juggins, 1993) used similar data generated by the same program COMPAS (Minchin, 1987), to examine the performance of CCA and various statistical procedures ill-adapted to deal with unimodal species responses. CCA was assessed as being robust to the relaxation of the assumptions and as being better than the statistical alternatives. In neither the work of Palmer nor ter Braak were alternative methods better adapted to a more complete ecological model of species responses such as non-metric multidimensional scaling using the Bray–Curtis coefficient (Faith et al., 1987) tested.

Guisan et al. (1999) provide a comparison of GLM with CCA. The GLM modelling included cubic polynomials and interaction terms. The authors conclude that the spatial distribution of individual species is better modelled by GLM than CCA. However, CCA may have some advantages for rare species with few positive observations. The combined use of information from other species in defining the underlying gradient helps to define the behaviour of the rare species. Ejrnaes (2000) also considers that DCA may have some advantages over CCA.

The simulated studies (Palmer, 1993; ter Braak et al., 1993) used reasonable vegetation data but the environmental data were unrealistic. The use of real data always runs the risk that none of the methods compared recovers the full truth. ‘Truth’ in real data sets must always remain uncertain (Austin et al., 1995). In any research some assumptions must be taken as given but should be questioned as soon as possible.

### 3.3. *Ecological theory tested*

Relatively few papers have attempted to test existing vegetation theory. Austin (1987) at-

tempted to test the propositions of Gauch and Whittaker (1972) about species patterns along gradients. He found that positive-skewed unimodal response curves were typical of eucalypt forest in southeastern Australia. However, tree species richness was found to increase along a temperature gradient (Austin, 1987). This was in contrast to Gauch and Whittaker (1972) who thought there was no consistent relationship of species richness with gradients from a subjective evaluation of their data. Other theoretical propositions could not be tested due to the confounding effect of the gradient in species richness. Note running means were used to describe the species response curves. Minchin (1989), in a more extensive analysis using 100 vascular plant species and statistical modelling, tested several hypotheses. He found that only 45% of species had unimodal, symmetric response curves. The modes of the ‘major’ species were randomly distributed rather than evenly distributed. Other hypotheses gave inconsistent results between different structural (growth form) species groups. Total species richness (alpha diversity) had a complex trend surface pattern in relation to the two complex indirect environmental gradients studied. However, each structural group of species had a response pattern in accord with the hypothesis that species richness shows a unimodal response along environmental gradients.

Two conclusions in addition to rejection of the bell-shaped curve hypothesis can be drawn from these studies. Changes in species richness along a gradient are inconsistent with assumptions about uniform distributions of species modes (ter Braak, 1986). Species richness is a collective property of vegetation, therefore, that needs to be incorporated into models (Margules et al., 1987; Huston, 1994; Pausas, 1994; Austin et al., 1996; Pausas and Austin, 2001). Secondly, an adequate model of community variation along environmental gradients must take into account differences in response patterns between species groups with different growth forms (Minchin, 1989).

Recently, others with more powerful statistical models have tested hypotheses about species response curves and the appropriateness of the CCA model. Bio et al. (1998) tested the species

response shapes of 156 species in relation to six environmental variables using GAM. Twenty-three percent of the 156 models contained exclusively sigmoid (linear) and Gaussian response curves. (Note linear responses arise when species distributions extend beyond the extremities of the gradient sampled.) The remaining 77% of models had at least one response function that was smooth and more complex. Ejrnaes (2000) examined also with GAM, the response functions of 146 species in relation to a single pH gradient. He found 20% of species with symmetric unimodal curves. In addition, Ejrnaes (2000) compared the performance of modelling with environmental predictors as opposed to using DCA axes as predictors. He avoided circularity by deriving DCA ordinations minus the species being modelled. Twenty-two species were modelled. Better predictive models were obtained with DCA axes for 19 of the species. The conclusion is drawn that the species are responding to the same fundamental environmental gradients. One may conclude other species are better estimators of the gradients than an arbitrary selection of environmental variables. Ejrnaes (2000) suggests that this may be a more important effect than deviations from the symmetric unimodal response (Austin and Gaywood, 1994). Vayssieres et al. (2000) also examined the shape of species responses during a comparison of GLM and CART methods using three Oak species. Two of the three species were fitted with models including cubic polynomials and interaction terms.

Guisan et al. (1999) in a comparison of GLM and CCA found polynomials and interaction terms were useful in modelling species distributions. They concluded that GLM gave better models for individual species but that CCA had some positive features. All the studies agree that symmetric unimodal responses are rare and skewed curves predominate. More detailed comparison is difficult because no two studies have used the same approach or similar types of vegetation data. Ejrnaes (2000) and Guisan et al. (1999) agree that CCA types of constrained regression have advantages but there is no consensus (Austin et al., 1994; Bio et al., 1998). Detailed analyses of species patterns along envi-

ronmental gradients using GLM, GAM or CART to describe the individual responses have not been made. Species packing models have been neglected in hypothesis testing so far. One exception is the work of Austin et al. (1994), Austin and Gaywood (1994), Austin (1999b). It was hypothesised that species responses along an environmental gradient would have a skewed shape with tails towards the mesic central portion of the gradient and modes with steep declines towards the extremes of the gradient. The process basis of this was that species limits towards the extremes were determined by physiological tolerances while internal limits were determined by competition (Austin, 1990). This was confirmed by GLM models applied to eucalypt species in forests of southeastern Australia (Austin et al., 1994; Austin and Gaywood, 1994). The generality of the result requires testing. Eucalypt forests in Australia are a unique vegetation type. All the canopy species belong to one genus (*sensu lato*). Although several environmental predictors were used in the successful models, the pattern was only found for the temperature gradient. The pattern applied to species with temperature optima within the range studied, i.e. temperate species; those species with optima at higher temperature were not included, i.e. subtropical species with truncated responses in the region.

This work has led to controversy. This concerned the estimation of the coefficients of the  $\beta$ -function  $((x-a)^\alpha, (b-x)^\gamma)$  when it is used to fit species response curves, particularly the dependence of the  $\alpha$  and  $\gamma$  values on  $a$  and  $b$  values (Oksanen, 1997). Austin and Nicholls (1997) responded demonstrating that the results depended critically on how data limits (i.e.  $a$  and  $b$  values) were set for modelling individual species. Oksanen and Minchin (2002) have now examined a wider range of possible models for the shape of species response curves. A significant issue remains. Should species whose response curves are clearly truncated, i.e. their range extends beyond the limits of the gradient sampled, be used in theoretical studies of response shape? There is then no data on which to base an estimate of the species limit. Failure to have an agreed data model is confusing tests of theory.

Progress in testing hypotheses concerning vegetation patterns has been slow. Progress in statistical methods, remote-sensing and GIS has outstripped that in ecological theory (Franklin et al., 2000). Even when the purpose is prediction, a better synthesis of theory will inform modelling and less biased predictions will result. A systematic approach to testing the ecological theories implicit in many of the current predictive modelling approaches is needed.

#### 4. Role of descriptive analytical ecology

Since a comprehensive theory of vegetation will not result from correlative studies of vegetation/environment relationships alone, what role is there for such descriptive studies? Any mechanistic process model of ecosystem dynamics should be consistent with a static, quantitative and rigorous description of the same ecosystem. Statistical habitat models offer a means of achieving that description. Discrepancies between process model and statistical model will demonstrate our lack of knowledge and may also indicate the way forward. Before progress can be made we must recognise current limitations.

There are too few studies to evaluate how best to measure either the vegetation or the environmental determinants of vegetation (Guisan, 2002; Guisan and Harrell, 2000; Guisan and Zimmermann, 2000; Franklin et al., 2000). We know little about the relative importance of different environmental and biotic variables in different biomes (Meentemeyer et al., 2001). Austin and Smith (1989) summarised the incompleteness of many conceptual frameworks for plant community ecology. There has been little progress in theoretical plant community ecology since, which is relevant to predicting the realised niche of a species (Austin and Gaywood, 1994; Austin, 1999a,b; Guisan and Zimmermann, 2000).

A regression study (Austin, 1971) provided some insight into the relationships between distal and proximal variables, indirect and direct gradients and plant competition when re-examined (Austin, 1979). Distal variables were shown to be important near the limits of a species distribution

while proximal variables assumed more importance as predictors under more optimal conditions for the species. The regression models and the importance of competition effects varied with plant community the species was occurring in. Initial regression modelling of the stem density per plot of *Eucalyptus roysii* in relation to simple indirect environmental predictors yielded negligible results (Fig. 4). Inclusion of curvilinear functions made little difference. A series of strategies were then introduced into the regression-building process based on ecological ideas. The data were stratified into different plant communities based on a numerical classification of the 261 sites using 352 species. *E. roysii* occurred in only four of the eight communities recognised. Limits were set on the data to exclude those sites clearly beyond the realised niche of the species (mean annual rainfall > 810 mm and altitude above 1100 m). More proximal and direct variables were estimated from the available environmental variables. A solar radiation index was calculated from slope, aspect and latitude, termed a derived variable in Fig. 4. Moisture stress indices were estimated from these derived variables using a water balance model. The most abundant eucalypt species in each community were then added to the regression models as potential competition variables. Finally, as the variance was found to be increasing with the mean, a weighted regression procedure was introduced. The increase in ecological realism increased the predictive success of the models, at least in terms of the adjusted  $r^2$  (Fig. 4, Austin, 1979). The importance of different steps varied with the community. An ecological interpretation was also possible for the models obtained for the individual communities. The model for the western dry sclerophyll forest dominated by *E. roysii*, contained curvilinear terms for various components of the water balance (Austin, 1971). These included moisture stress for the main drought period, available soil moisture and mean annual rainfall. No competition terms were significant. The model for the eastern dry sclerophyll forest where *E. roysii* was a subordinate species, contained competition terms for two eucalypt species. Other terms were moisture stress index plus simple linear terms for rainfall, pH and solar radiation index.

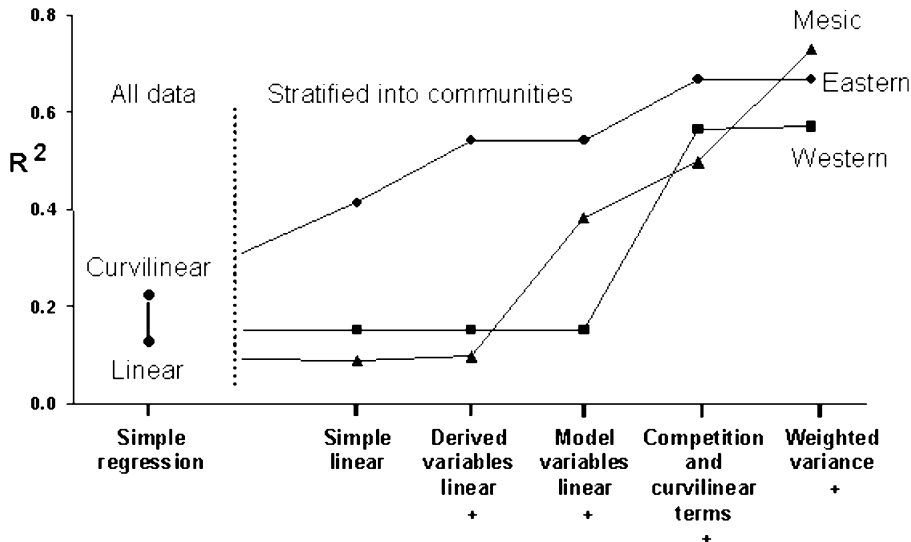


Fig. 4. Relative success of exploratory regression analysis predicting *E. rossii* stem density in dry sclerophyll forest in south-eastern Australia. Success as measured by  $r^2$  increases after stratification and with increasingly relevant ecological and statistical models. (Redrawn from Austin, 1979, with permission of International Co-operative Publishing company.)

Distal variables such as rainfall are likely to be the limiting factor in any water balance model when a species is at the limits of its distribution. It is under these same conditions that competition from other species is most likely to have an influence. The models are consistent with ecological expectations (Austin, 1971). The importance of distal and proximal variables varies with the ecological context. Where gradients are steep and environments are extreme, simple distal variables will be as successful in modelling as environmental process based models such as moisture stress. Where environment is changing slowly and a species is occupying its optimal realised niche, then proximal variables will be more successful predictors. The success of different habitat modelling strategies will be context-sensitive. A similar exercise on the same data set with a different species, *E. mannifera*, but using GLM, gave similar results (Austin and Cunningham, 1981). A similar modelling strategy needs to be tested on data sets from different biomes.

## 5. Conclusion

There is an intimate relationship between theory and method. In plant community ecology that relationship is often neglected, as it is in other branches of ecology (Austin, 1999a,b; Scott et al., 2002). When vegetation studies are interfaced with statistics, the problems of the relationship of theory and methodology in two distinct disciplines have to be integrated.

The studies reviewed here reveal two areas where ecological theory is inadequate and methods suspect. These are patterns of species composition and estimation of environmental predictors.

### 5.1. Patterns of vegetation composition

There is no synthesis of theories regarding the expected pattern of vegetation composition in relation to environment (Gauch and Whittaker, 1972; Grime, 1979; Austin and Smith, 1989; Keddy, 1989; Huston, 1994). The collective property, species richness is accorded great importance

in experimental plant community ecology (Naeem et al., 1996; Tilman et al., 1996; Austin, 1999b). Yet its impact on vegetation analysis methods such as CCA remains to be examined. Clear patterns of response to environmental gradients at a regional scale have been demonstrated (Margules et al., 1987; Minchin, 1989; Pausas, 1994; Austin et al., 1996; Leathwick et al., 1998). These patterns do not relate in any simple fashion to individual species responses, though patterns of richness do depend on growth form types (Minchin, 1989).

### 5.2. Estimation of environmental predictors

The second area is in the development of measures for environmental predictors and the extent to which knowledge of environmental process models should be incorporated (Austin, 1979; Guisan and Zimmermann, 2000; Franklin et al., 2000). The studies of Austin (1971, 1979), Austin and Cunningham (1981), Leathwick (1995, 1998), Leathwick et al. (1996, 1998), Leathwick and Austin (2001) suggest that process models and even variables for competition from other species have a role to play in static statistical models. The circumstances when they should be used and the strategy for introducing such variables into the modelling procedure remain uncertain. Statistical theory and methods for spatial autocorrelation and regression smoothing will have a big impact in how we address the ecological issues. The relative neglect of interaction terms in recent vegetation modelling is a current limitation to progressing beyond a one-dimensional theory of vegetation composition. The apparent power of CART techniques (Austin et al., 1995; O'Connor et al., 1996; Iverson and Pasad, 1998; Franklin, 1998) to detect interactions is only one example of the need for new strategies for predictive habitat modelling if they are to address ecological theory as well as predicting species distributions for immediate practical conservation purposes.

### 5.3. Data model and statistical methods

Current practice with respect to the data model is to emphasise the use of stratified sampling (Austin and Heyligers, 1989; Guisan and Zimmer-

mann, 2000; Iverson and Pasad, 1998; Franklin et al., 2002; Cawsey et al., 2002), and to use either presence–absence data or cover/abundance values. Guisan and Harrell (2000) indicate that better statistical models (ordinal regression) are required for use with cover/abundance data. Preferred statistical models at present are GAM (Yee and Mitchell, 1991; Leathwick, 1995), or CART (Franklin, 1998). Use of these methods to define the functions to be incorporated into a parametric GLM needs to be considered. GAM may produce a complex function where an equivalent parametric function may capture most of the same variation and have a more reasonable ecological explanation. CART by its nature must produce discontinuities when a continuous function may be more realistic. Best practice may be to analyse data with both GAM and CART using the results to construct as far as possible a GLM with ecologically rational functions. It is clear, however, that best practice will need to incorporate spatial autocorrelation (Leathwick, 1998 for example) probably by means of an autologistic approach (Augustin et al., 1996). Closer collaboration with statisticians is likely to see many new methods being examined (T. Hastie pers. com.). There are further complications from an ecological theory viewpoint that need to be examined in future. The role of ecotypic variation within a species influencing species response curves remains to be examined. The modelling of species reproductive potential in response to environment is in its infancy (Heegaard, 2001). The inconsistency between the realised niche and source-sink concepts of species distributions (Pulliam, 1988; Heegaard, 2001) has to be resolved for plants. The upper and lower limits in relation to temperature will not be determined by the same proximal variables so species statistical response models are unlikely to have simple unimodal functions. The estimation of spatial autocorrelation effects may help in detecting this phenomenon and other biotic processes influencing distributions. It remains to be seen to what extent vegetation is in equilibrium with environment and the implications of this for static analysis. Analytical description of species distributions and the collective properties of vegetation can address these issues. Ultimately, theory, phy-

biological process knowledge, and dynamic modelling must be tested for explanatory power against quantitative description of the natural patterns.

## Acknowledgements

I thank J. Reid, A.O. Nicholls, A. Guisan and R.J. O'Connor for comments on the manuscript and to S. Marsden for help with its preparation.

## References

- Aspinall, R., Veitch, N., 1993. Habitat mapping from satellite imagery and wildlife survey data using a Bayesian modelling procedure in a GIS. *P E & R S* 59, 537–543.
- Augustin, N., Muggleston, M., Buckland, S., 1996. An autologistic model for spatial distribution of wildlife. *J. Appl. Ecol.* 33, 339–347.
- Austin, M.P., 1971. Role of regression analysis in plant ecology. *Proc. Ecol. Soc. Aust.* 6, 63–75.
- Austin, M.P., 1979. Current approaches to the non-linearity problems in vegetation analysis. In: Patil, G.P., Rosenzweig, M.L. (Eds.), *Contemporary Quantitative Ecology and Related Ecometrics*. International Co-operative Publishing House, Fairland, MD, pp. 197–210.
- Austin, M.P., 1980. Searching for a model for use in vegetation analysis. *Vegetatio* 42, 11–21.
- Austin, M.P., 1985. Continuum concept, ordination methods and niche theory. *Ann. 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., 1990. Community theory and competition in vegetation. In: Grace, J.B., Tilman, D. (Eds.), *Perspectives on Plant Competition*. Academic Press, California, pp. 215–238.
- Austin, M.P., 1991. Vegetation: data collection and analysis. In: C.R. Margules, M.P. Austin (Eds.) *Nature Conservation: Cost-Effective biological surveys and data analysis*, Proceedings of CONCOM Workshop, pp. 37–41.
- Austin, M.P., 1992. Modelling the environmental niche of plants: implications for plant community response to elevated CO<sub>2</sub> levels. *Aust. J. Botany* 40, 615–630.
- Austin, M.P., 1998. An ecological perspective on biodiversity investigations: examples from Australian eucalypt forests. *Ann. Mo. Bot Gard.* 85, 2–17.
- Austin, M.P., 1999a. A silent clash of paradigms: some inconsistencies in community ecology. *Oikos* 86, 170–178.
- Austin, M.P., 1999b. The potential contribution of vegetation ecology to biodiversity research. *Ecography* 22, 465–484.
- Austin, M.P., 2002. Case studies of the use of environmental gradients in vegetation and fauna modelling: theory and practice in Australia and New Zealand. In: Scott, J.M., Heglund, P.J., Samson, F., Haufler, J., Morrison, M., Raphael, M., Wall, B. (Eds.), *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island Press, Covelo, CA, pp. 73–82.
- Austin, M.P., Cunningham, R.B., 1981. Observational analysis of environmental gradients. *Proc. Ecol. Soc. Aust.* 11, 109–119.
- 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., Gaywood, M.J., 1994. Current problems of environmental gradients and species response curves in relation to continuum theory. *J. Veg. Sci.* 5, 473–482.
- Austin, M.P., Nicholls, A.O., 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.
- Austin, M.P., Smith, T.M., 1989. A new model for the continuum concept. *Vegetatio* 83, 35–47.
- Austin, M.P., Nicholls, A.O., Doherty, M.D., Meyers, J.A., 1994. Determining species response functions to an environmental gradient by means of a  $\beta$ -function. *J. Veg. Sci.* 5, 215–228.
- Austin, M.P., Meyers, J.A., Belbin, I., Doherty, M.D., 1995. Modelling of landscape patterns and processes using biological data. Subproject 5: simulated data case study. Consultancy report for ERIN, CSIRO Wildlife and Ecology, Canberra.
- Austin, M.P., Pausas, J.G., Nicholls, A.O., 1996. Patterns of tree species richness in relation to environment in south-eastern New South Wales, Australia. *Aust. J. Ecol.* 21, 154–164.
- Bayliss, G.T.S., 1980. Mycorrhizas and the spread of beech. *New Zealand J. Ecol.* 3, 151–153.
- Begon, M., Harper, J.L., Townsend, C.R., 1990. *Ecology, Individuals, Populations and Communities*, second ed.. Blackwell Scientific Publications, Boston, p. 945.
- Birks, H.J.B., Frey, D.G., Deevey, E.S., 1998. Review #1 numerical tools in palaeolimnology-progress, potentialities, and problems. *J. Palaeolimnol.* 20, 307–332.
- Birks, H.J.B., Austin, H.A., 1992. An Annotated Bibliography of Canonical Correspondence Analysis and Related Constrained Ordination Methods 1986–1991. Botanical Institute, University of Bergen, Bergen, Norway.
- Bio, A.M.F., Alkemade, R., Barendregt, A., 1998. Determining alternative models for vegetation response analysis: a non-parametric approach. *J. Veg. Sci.* 9, 5–16.
- Boyce, M.S., Vernier, P.R., Nielsen, S.E., Schmiegelow, F.K.A., 2002. Evaluating resource selection functions. *Ecol. Model.* 157 (2–3), 279–298.
- Box, G.E.P., 1966. Use and abuse of regression. *Technometrics* 8, 625–629.
- Braun-Blanquet, J., 1964. *Pflanzensoziologie*, third ed.. Springer, Wien.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. *Classification and Regression Trees*. Wadsworth International Group, Belmont, CA, USA, p. 358.

- Brunet, J., von Oheimb, G., Diekmann, M., 2000. Factors influencing vegetation gradients across ancient–recent woodland borderlines in southern Sweden. *J. Veg. Sci.* 11, 515–524.
- Brzeziecki, B., 1987. Analysis of vegetation–environment relationships using a simultaneous equations model. *Vegetatio* 71, 175–184.
- Cawsey, E.M., Austin, M.P., Baker, B.L., 2002. Regional vegetation mapping in Australia: a case study in the practical use of statistical modelling. *Biodivers. Conserv.*, in press.
- Cornell, H.V., Karlson, R.H., 1997. Local and regional processes as controls of species richness. In: Tilman, D., Kareiva, P. (Eds.), *Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions*. Princeton University Press, Princeton, pp. 250–268.
- Cornell, H.V., Lawton, J.H., 1992. Species interactions, local and regional processes, and limits to the richness of ecological communities: a theoretical perspective. *J. Anim. Ecol.* 61, 1–12.
- Currie, D.J., 1991. Energy and large-scale patterns of animal and plant-species richness. *Am. Naturalist* 137, 27–49.
- Elith, J., Burgman, M.A., 2002. Habitat models for PVA. In: Bringham, C.A., Schwanz, M.W. (Eds.), *Population Viability in Plants*. Springer.
- Ejrnæs, R., 2000. Can we trust gradients extracted by detrended correspondence analysis. *J. Veg. Sci.* 11, 565–572.
- Ejrnæs, R., Bruun, H.H., 2000. Gradient analysis of dry grassland vegetation in Denmark. *J. Veg. Sci.* 11, 573–584.
- Faith, D.P., Minchin, P.R., Belbin, I., 1987. Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* 69, 57–68.
- Fewster, R.M., Buckland, S.T., Siriwardena, G.M., Baillie, S.R., Wilson, J.D., 2000. Analysis of population trends for farmland birds using generalized additive models. *Ecology* 81, 1970–1984.
- Fitzgerald, R.W., Lees, B.G., 1992. The application of neural networks to the floristic classification of remote sensing and GIS data in complex terrain. In: *Proceedings of the 17th Congress of the International Society for Photogrammetry and Remote Sensing*, Washington, USA.
- Franklin, J., 1995. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Prog. Phys. Geogr.* 19, 474–499.
- Franklin, J., 1998. Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. *J. Veg. Sci.* 9, 733–748.
- Franklin, J., McCullough, P., Gray, C., 2000. Terrain variables used for predictive mapping of vegetation communities in southern California. In: Wilson, J.P., Gallant, J.C. (Eds.), *Terrain Analysis: Principles and Applications*. Wiley.
- Franklin, J., Keeler-Wolf, T., Thomas, K.A., Shaari, D.A., Stine, P.A., Michaelsen, J., Miller, J., 2002. Stratified sampling for field survey of environmental gradients in the Mojave desert ecoregion. In: Millington, A., Walsh, S., Osborne, P. (Eds.), *GTS and Remote-Sensing Applications in Biogeography and Ecology*. Kluwer Academic Publishers.
- Gauch, H.G., Whittaker, R.H., 1972. Coenocline simulation. *Ecology* 53, 446–451.
- Giller, J., 1984. *Community Structure and the Niche*. Chapman & Hall, London, p. 176.
- Gleason, H.A., 1926. The individualistic concept of the plant association. *Bull. Torrey Botanical Club* 53, 7–26.
- Grime, J.P., 1979. *Plant Strategies and Vegetation Processes*. Wiley, New York.
- Guisan, A., 2002. Semi-quantitative response models for predicting the spatial distribution of plant species. In: Scott, J.M., Heglund, P.J., Samson, F., Haufler, J., Morrison, M., Raphael, M., Wall, B. (Eds.), *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island Press, Covelo, CA, pp. 315–326.
- Guisan, A., Harrell, R.E., 2000. Ordinal response regression models in ecology. *J. Veg. Sci.* 11, 617–626.
- Guisan, A., Theurillat, J.-P., 2000. Equilibrium modelling of alpine plant distribution: how far can we go. *Phytocoenologia* 30, 353–384.
- Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. *Ecol. Model.* 135, 147–186.
- Guisan, A., Weiss, S.B., Weiss, A.D., 1999. GLM versus CCA spatial modelling of plant species distribution. *Plant Ecol.* 143, 107–122.
- Gumpertz, M.L., Graham, J.M., Ristaino, J.B., 1997. Autologistic model of spatial pattern of phytophthora epidemic in Bell Pepper: effects of soil variables on disease presence. *J. Agric. Biol. Environ. Stat.* 2, 131–156.
- Hastie, T., Tibshirani, R., 1990. *Generalised Additive Models*. Chapman & Hall, London.
- Heegaard, E., 2001. Environmental relationships of perichaetial and sporophyte production in *Andreaea* spp. in Western Norway. *J. Bryol.* 23, 97–108.
- Huston, M.A., 1994. *Biological Diversity: The Coexistence of Species on Changing Landscapes*. Cambridge University Press, Cambridge.
- Iverson, L.R., Pasad, A.M., 1998. Predicting abundance of 80 tree species following climate change in the eastern United States. *Ecol. Monogr.* 68, 465–485.
- Keddy, P.A., 1982. Population ecology on an environmental gradient: *Cattle edentula* on a sand dune. *Oecologia* 52, 348–355.
- Keddy, P.A., 1989. *Competition*. Chapman & Hall, London.
- Krebs, C.J., 1994. *Ecology: The experimental Analysis of the Distribution of Distribution and Abundance*, fourth ed.. Harper & Row, New York.
- Leathwick, J.R., 1995. Climatic relationships of some New Zealand forest tree species. *J. Veg. Sci.* 6, 237–248.
- Leathwick, J.R., 1998. Are New Zealand's *Nothofagus* species in equilibrium with their environment. *J. Veg. Sci.* 9, 719–732.
- Leathwick, J.R., 2002. Intra-generic competition among *Nothofagus* in New Zealand's old-growth forests. *Biodivers. Conserv.*, in press.



- Leathwick, J.R., Austin, M., 2001. Competitive interactions between tree species in New Zealand old-growth indigenous forests. *Ecology* 82, 2560–2573.
- Leathwick, J.R., Mitchell, N.D., 1992. Forest pattern, climate and vulcanism in central North Island, New Zealand. *J. Veg. Sci.* 3, 603–616.
- Leathwick, J.R., Whitehead, D., McLeod, M., 1996. Predicting changes in the composition of New Zealand's indigenous forests in response to global warming: a modelling approach. *Environ. Software* 11, 81–90.
- Leathwick, J.R., Burns, B.R., Clarkson, B.D., 1998. Environmental correlates of tree alpha-diversity in New Zealand primary forests. *Ecography* 21, 235–246.
- Leps, J., Michalek, J., Rauch, O., Uhlik, P., 2000. Early succession on plots with the upper soil horizon removed. *J. Veg. Sci.* 11, 259–264.
- McCullagh, P., Nelder, J.A., 1989. *Generalized Linear Models*, second ed.. Chapman & Hall, London.
- Malanson, G.P., Westman, W.E., Van, Y.-L., 1992. Realized versus fundamental niche functions in a model of chaparral response to climatic change. *Ecol. Model.* 64, 261–277.
- Margules, C.R., Nicholls, A.O., Austin, M.P., 1987. Diversity of *Eucalyptus* species predicted by a multi variables environmental gradient. *Oecologia* 71, 229–232.
- McGlone, M.S., Mildenhall, D.C., Pole, M.S., 1996. History and palaeoecology of New Zealand *Nolkofoagus* forests. In: Veblin, T.T., Hill, R.S., Read, J. (Eds.), *The Ecology and Biogeography of Nothofagus Forests*. Yale University Press, New Haven, pp. 83–130.
- Meentemeyer, R.K., Moody, A., Franklin, J., 2001. Landscape-scale patterns of shrub-species abundance in California chaparral: the role of topographically mediated resource gradients. *Plant Ecol.* 156, 19–41.
- Minchin, P.R., 1987. Simulation of multidimensional community patterns: towards a comprehensive model. *Vegetatio* 71, 145–156.
- Minchin, P.R., 1989. Montane vegetation of the Mt. Field Massif, Tasmania: a test of some hypotheses about properties of community patterns. *Vegetatio* 83, 97–110.
- Mueller-Dombois, D., Ellenberg, H., 1974. *Aims and Methods of Vegetation Ecology*. Wiley, New York, p. 547.
- Naem, S., Hakansson, K., Lawton, J.H., Crawley, M.J., Thompson, L.J., 1996. Biodiversity and plant productivity in a model assemblage of plant species. *Oikos* 76, 259–264.
- Naem, S., Thompson, L.J., Lawler, S.P., Lawton, J.H., Woodfin, R.M., 1994. Declining biodiversity can alter the performance of ecosystems. *Nature* 368, 734–737.
- O'Connor, R.J., 2001. The conceptual basis of species distribution modelling: time for paradigm shift. In: Scott, J.M., Heglund, P.J., Samson, F., Haufler, J., Morrison, M., Raphael, M., Wall, B. (Eds.), *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island Press, Covelo, CA, pp. 25–33.
- O'Connor, R.J., Jones, M.T., White, D., Hunsaker, C., Loveland, T., Jones, B., Preston, E., 1996. Spatial partitioning of the environmental correlates of avian biodiversity in the lower United States. *Biodivers. Lett.* 3, 97–110.
- Oksanen, J., 1997. Why the beta-function cannot be used to estimate skewness of species responses. *J. Veg. Sci.* 8, 147–152.
- Oksanen, J., Minchin, P.R., 2002. Continuum theory revisited: what shape are species responses along ecological gradients? *Ecol. Model.* 157 (213), 119–129.
- Palmer, M.W., 1993. Putting things in even better order: the advantages of canonical correspondence analysis. *Ecology* 74, 2215–2230.
- Partel, M., Zobel, M., Zobel, K., van der Maarel, E., 1996. The species pool and its relation to species richness: evidence from Estonian plant communities. *Oikos* 75, 111–117.
- Pausas, J.G., 1994. Species richness patterns in the understory of Pyrenean *Pinus sylvestris* forests. *J. Veg. Sci.* 5, 517–524.
- Pausas, J.G., Austin, M.P., 2001. Patterns of plant species richness in relation to different environments: an appraisal. *J. Veg. Sci.* 12, 153–166.
- Pereira, J.M.C., Itami, R.M., 1991. GIS-based habitat modeling using logistic multiple regression: a study of the MT. Graham Red Squirrel. *PERS* 57, 1475–1486.
- Pulliam, H.R., 1988. Sources, sinks, and population regulation. *Am. Nat.* 132, 652–661.
- Ricklefs, R.B., Schluter, D. (Eds.), *Species Diversity in Ecological Communities: Historical and Geographical Perspectives*. University Chicago Press, Chicago 1993.
- Rosenzweig, M.L., 1995. *Species Diversity in Space and Time*. Cambridge University Press, Cambridge, UK.
- Scott, J.M., Heglund, P.J., Samson, F., Haufler, J., Morrison, M., Raphael, M., Wall, B., 2002. *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island Press, Covelo, CA, pp. 868.
- Sokal, R.R., Rohlf, F.J., 1981. *Biometry The principles and Practice of Statistics in Biological Research*, second ed.. W.H. Freeman, San Francisco.
- ter Braak, C.J.F., 1985. Correspondence analysis of incidence and abundance data: properties in terms of a unimodal response model. *Biometrics* 41, 859–873.
- ter Braak, C.J.F., 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* 67, 1167–1179.
- ter Braak, C.J.F., 1987. The analysis of vegetation–environment relationships by canonical correspondence analysis. *Vegetatio* 69, 69–77.
- ter Braak, C.J.F., Juggins, S., 1993. Weighted averaging partial least squares regression (WA-PLS): an improved method for reconstructing environmental variables from species assemblages. *Hydrobiologia* 269/270, 485–502.
- ter Braak, C.J.F., Verdonschot, P.F.M., 1995. Canonical correspondence analysis and related multivariate methods in aquatic ecology. *Aquat. Sci.* 57, 253–287.
- ter Braak, C.J.F., Juggins, S., Birks, H.J.B., van der Voel, H., 1993. Weighted averaging partial least squares regression (WA-PLS): definition and comparison with other methods for species–environment calibration. In: Patil, G.P., Rao, C.R. (Eds.), *Multivariate Environmental Statistics, North-Holland Series in Statistics and Probability*, vol. 6, pp. 525–560.

- Tilman, D., Wedin, D., Knops, J., 1996. Productivity and sustainability influenced by biodiversity in grassland ecosystems. *Nature* 379, 718–720.
- Tilman, D., Knops, J., Wedin, D., Reich, P., Ritchie, M., Siemann, E., 1997. The influence of functional diversity and composition on ecosystem processes. *Science* 277, 1300–1302.
- Vayssières, M.P., Plant, R.E., Allen-Diaz, B.H., 2000. Classification trees: an alternative non-parametric approach for predicting species distributions. *J. Veg. Sci.* 11, 679–694.
- Watkinson, A.R., 1985. On the abundance of species along an environmental gradient. *J. Ecol.* 73, 569–578.
- Watkinson, A.R., Sutherland, W.J., 1995. Sources, sinks and pseudo-sinks. *J. Anim. Ecol.* 64, 126–130.
- Weir, I.S., Pettit, A.N., 2000. Binary probability maps using a hidden conditional autoregressive Gaussian process with an application to Finnish common toad data. *Appl. Stat.* 49, 473–484.
- Westhoff, V., van der Maarel, E., 1978. The Braun–Blanquet approach. In: Whittaker, R.H. (Ed.), *Classification of Plant Communities*, second ed.. Junk, The Hague, pp. 287–399.
- Wu, H., Huffer, F.W., 1997. Modelling the distribution the distribution of plant species using the autologistic regression model. *Environ. Ecol. Stat.* 4, 49–64.
- Yce, T.W., Mitchell, N.D., 1991. Generalized additive models in plant ecology. *J. Veg. Sci.* 2, 587–602.