Species Distribution Models: Ecological Explanation and Prediction Across Space and Time

Jane Elith[1] and John R. Leathwick[2]

[1]School of Botany, The University of Melbourne, Victoria 3010, Australia;
email: j.elith@unimelb.edu.au

[2]National Institute of Water and Atmospheric Research, Hamilton, New Zealand;
email: j.leathwick@niwa.co.nz

Annu. Rev. Ecol. Evol. Syst. 2009. 40:677–97

The *Annual Review of Ecology, Evolution, and Systematics* is online at ecolsys.annualreviews.org

This article's doi:
10.1146/annurev.ecolsys.110308.120159

1543-592X/09/1201-0677$20.00

Key Words

climate change, invasions, niche, predict, presence-only, spatial

Abstract

Species distribution models (SDMs) are numerical tools that combine observations of species occurrence or abundance with environmental estimates. They are used to gain ecological and evolutionary insights and to predict distributions across landscapes, sometimes requiring extrapolation in space and time. SDMs are now widely used across terrestrial, freshwater, and marine realms. Differences in methods between disciplines reflect both differences in species mobility and in "established use." Model realism and robustness is influenced by selection of relevant predictors and modeling method, consideration of scale, how the interplay between environmental and geographic factors is handled, and the extent of extrapolation. Current linkages between SDM practice and ecological theory are often weak, hindering progress. Remaining challenges include: improvement of methods for modeling presence-only data and for model selection and evaluation; accounting for biotic interactions; and assessing model uncertainty.

INTRODUCTION

Throughout the centuries humans have observed and recorded consistent relationships between species distributions and the physical environment. Whilst early scientific writings were largely qualitative (Grinnell 1904), numerical models are now widely used both for describing patterns and making predictions. These numerical techniques support a rich diversity of applications, arguably with varying degrees of success. Published examples indicate that species distribution models (SDMs) can perform well in characterizing the natural distributions of species (within their current range), particularly when well-designed survey data and functionally relevant predictors are analyzed with an appropriately specified model. In such a setting, models can provide useful ecological insight and strong predictive capability. By contrast, applications that fit models for species not substantially in equilibrium with their environment, that extrapolate in time or space, and/or use inadequate data are much more challenging, and results are more equivocal.

Our aim is to review the history and current status of the SDM literature, exploring applications spanning biological realms and scientific disciplines. We define an SDM as a model that relates species distribution data (occurrence or abundance at known locations) with information on the environmental and/or spatial characteristics of those locations (for key steps, see Sidebar, Basics of Species Distribution Modeling). The model can be used to provide understanding and/or to predict the species' distribution across a landscape. Names for such models vary widely. What we term SDMs have also been called (sometimes with different emphases and meanings): bioclimatic models, climate envelopes, ecological niche models (ENMs), habitat models, resource selection functions (RSFs), range maps, and—more loosely—correlative models or spatial models. We include these, but exclude models that are mechanistic or process-based (see Kearney & Porter 2009 for a review), or that predict community-level features such as community composition and species turnover or richness (see Ferrier & Guisan 2006 for a review).

Reviews of SDM literature include those of Guisan & Zimmermann (2000), Stauffer (2002), Guisan & Thuiller (2005), Richards et al. (2007), and Schröder (2008). Several books have either been recently published or are in preparation (Franklin 2009; A.T. Peterson & A. Guisan, personal communication). Instructional texts and training opportunities in species modeling are now available, including online texts (Pearson 2007) and university courses and workshops.

In light of these resources, we provide only a brief review of the technical aspects of SDMs and do not give methodological advice, concentrating instead on historical and cross-disciplinary features. In particular, we probe the motivations and concepts inherent in different approaches, attempting to identify commonalities that are widely relevant, regardless of discipline boundaries. We explore the diverse uses of SDMs (across environments, spatial and temporal scales, and modeling techniques), including earlier emphases on understanding ecological relationships

BASICS OF SPECIES DISTRIBUTION MODELING

Key steps in good modeling practice include the following: gathering relevant data; assessing its adequacy (the accuracy and comprehensiveness of the species data; the relevance and completeness of the predictors); deciding how to deal with correlated predictor variables; selecting an appropriate modeling algorithm; fitting the model to the training data; evaluating the model including the realism of fitted response functions, the model's fit to data, characteristics of residuals, and predictive performance on test data; mapping predictions to geographic space; selecting a threshold if continuous predictions need reduction to a binary map; and iterating the process to improve the model in light of knowledge gained throughout the process (Elith & Leathwick 2009).

and the more recent focus on prediction. Finally, we identify and examine several emerging issues. Our limit of 120 references means that many interesting and relevant pieces of work inform our review but are not explicitly mentioned, so we also provide a **Supplemental Literature Cited** (follow the **Supplemental Material link** from the Annual Reviews home page at **http://www.annualreviews.org**) for download, listing useful papers for each topic.

Supplemental Material

THE SPECIES MODELING LANDSCAPE: ITS DEVELOPMENT AND DIVERSITY

Conceptual and Technical Underpinnings

Broadly speaking, contemporary SDMs combine concepts from ecological and natural history traditions with more recent developments in statistics and information technology. The ecological roots of SDMs belong in those early studies that described biological patterns in terms of their relationships with geographical and/or environmental gradients (e.g., Grinnell 1904, Murray 1866, Schimper 1903). Moreover, research that highlighted the individualistic responses of species to their environment (e.g., for vegetation, see Whittaker 1956; and for birds, see MacArthur 1958) provided the strong conceptual argument for modeling individual species rather than communities.

Modern quantitative modeling and mapping of species distributions emerged when two parallel streams of research activity converged. On the one hand, field-based ecological studies of species-habitat associations, at first reliant largely on linear multiple regression and discriminant function analyses (Capen 1981, Stauffer 2002), benefitted from new regression methods that provided coherent treatments for the error distributions of presence-absence and abundance data. Generalized linear models (GLMs) enabled pioneering regression-based SDMs that had much more sophistication and realism than was possible earlier (e.g., see Austin's work in 1970s and 1980s, cited in Austin 1985). The key structural features of GLMs (non-normal error distributions, additive terms, nonlinear fitted functions) continue to be useful and are part of many current methods including RSFs (Manly et al. 2002) and maximum entropy models (MaxEnt; Phillips et al. 2006).

In parallel, rapid methodological advances in physical geography provided new data and information systems. New methods allowed robust and detailed preparation of digital models of the Earth's surface elevation, interpolation of climate parameters, and remote sensing of surface conditions in both marine and terrestrial environments (see **Supplemental Literature Cited**). These greatly enhanced SDM capabilities by providing estimates of environmental conditions across entire landscapes, including retrospectively at surveyed locations. Alongside these advances, the development of geographic information systems (GIS) provided important tools for storing and manipulating both species records and environmental data (see Foody 2008; and Swenson 2008, who include accessible introductions to GIS). The gains are easily taken for granted, but stand in stark contrast to the resources available to early ecologists who usually only had simple measurements of location (e.g., latitude, longitude, and elevation or depth), and sometimes of local site conditions (e.g., slope, drainage, geology).

Early approaches to modeling species distributions within GIS used simple geographic envelopes, convex hulls, and environmental matching (e.g., Nix 1986; and see Section below, Methods for Modeling). SDMs as we think of them today emerged when the new statistical methods from field-based habitat studies were linked with GIS-based environmental layers. In one of the earliest applications of this integrated approach, Ferrier (1984, cited in Ferrier et al. 2002) applied GLMs (logistic regression) to predict the distribution of the Rufous scrub-bird using known locality records for the species, and remotely mapped and modeled environmental variables.

Models across Terrestrial, Freshwater, and Marine Environments

Spatial autocorrelation: when the values of variables sampled at nearby locations are not independent from each other

Species distributions have been modeled for terrestrial, freshwater and marine environments, and across species from many biological groups (see **Supplemental Literature Cited**). Terrestrial vascular plant analyses were prevalent in early years and are still common, along with studies of terrestrial animals (including invertebrates); marine and freshwater applications were relatively rare until the past 5–10 years, and soil-based organisms are still only infrequently modeled.

SDMs from these diverse fields display commonalities and contrasts, with differences in mobility between species prompting some major differences in modeling approach. When a species is sessile it is relatively easy to characterize its environment, even including the wider influence of landscape (e.g., the water flowing into a site can be modeled using topographic information). By contrast, mobile species tend to intermittently use resources that are patchily distributed across a landscape. Defining the environments sampled by such species at any given location can be challenging, particularly for some combinations of mobility and life-history characteristics. Models for mobile species with small home ranges are often fitted using methods similar to those for sessile organisms, perhaps with focal predictors summarizing information from the near-neighboring landscape (Ferrier et al. 2002). In contrast, models for highly mobile species (e.g., diadromous fish) need to include movement or access-related descriptors (e.g., stream-based distance to coast; Leathwick et al. 2008). RSFs or related techniques are useful for species where the important distinction is between locations that are "available" (can be reached by the animal, used or not) versus those that are "used" (for example, habitat selection studies for birds; Jones 2001).

Detection of mobile species can be problematic. In aquatic studies, observations are often treated as probabilities of capture and analyzed using similar methods as for sessile species, sometimes including temporal predictors to accommodate seasonal variation in catchability/presence (Venables & Dichmont 2004). Alternatively, specialized modeling techniques have been developed to account for imperfect detection (e.g., MacKenzie et al. 2002, Royle et al. 2004).

Historic differences in the way data are collected also create different emphases across disciplines. Plant quadrats are usually regarded as statistically independent samples provided they are sufficiently geographically separated. Continuous tow sampling is used for some marine organisms, resulting in loss of independence between samples located along the same tow. Similar problems exist for terrestrial transect samples and for samples from contiguous stream reaches. Such data have prompted use of mixed models or other methods for dealing with pseudoreplication and spatial autocorrelation (Dormann et al. 2007, and **Supplemental Literature Cited**).

Spatial Scale

Scale is relevant to the distributions of both species and environments, and comprises both grain and extent. The extent (or domain) usually reflects the purpose of the analysis. For instance, macroecological and global change studies tend to be continental to global in scope (e.g., Araújo & New 2007), whereas studies targeting detailed ecological understanding or conservation planning tend toward local to regional extents (Fleishman et al. 2001, Ferrier et al. 2002). Grain usually describes properties of the data or analysis—often the predictor variables and their grid cell size or polygon size, but also the spatial accuracy and precision of the species records (Dungan et al. 2002, Tobalske 2002). Grain should be consistent with the information content of the data, though in practice this is not always feasible, e.g., grids sometimes have to be defined at finer resolutions than the underlying data for consistency across predictors. A number of researchers have addressed the implications of using coarse- versus fine-scale data in SDMs (e.g., Ferrier & Watson 1997 and **Supplemental Literature Cited**), generally indicating that effects depend on the spatial accuracy of the data, characteristics of the terrain and species, and the intended application.

Conceptually there is no single natural scale at which ecological patterns should be studied (Levin 1992). Rather, the appropriate scale is dictated by the study goals, the system, and available data. Some species modelers emphasize notions of hierarchy in conceptualizing the influences of environment on species distributions (Allen & Starr 1982, Cushman & McGarigal 2002, Pearson & Dawson 2003). In terrestrial systems climate dominates distributions at the global scale (coarsest grain, largest extent), whereas at meso- and toposcales (a few to hundreds of kilometers) topography and rock type create the finer-scale variations in climate, nutrient availability, and water flows that influence species (e.g., Mackey & Lindenmayer 2001). Similarly, in freshwater ecosystems, hierarchical scales from watersheds to reaches to microhabitats all affect distributions (e.g., Poff 1997). Alternatively, scale can be considered from the species' viewpoint using the concept of selection orders (selection of microsite, patch, home range, population block, and geographic range) and focusing on the ways in which mobile animals interact with the spatial arrangement of environments (Addicott et al. 1987).

Although these are long-standing concepts, there is as yet little consensus on how to deal with scale disparities when fitting SDMs. Several methods, mostly from landscape ecology, focus on describing scales of pattern in ecological data. These include lacunarity, spectral analysis, and wavelet-coefficient regression (Saunders et al. 2005 and **Supplemental Literature Cited**). They provide useful tools for evaluating the inherent structure in data but their use for prediction seems underdeveloped. More commonly, analysts impose scales through data choice or model structure. Many do this unconsciously, using predictors likely to both vary and have effects on biota at markedly different spatial scales, but without explicit testing or discussion of the effect that this has on their results. Some deliberately construct a set of scale-dependent predictors to represent factors affecting the distribution of the target species at more than one spatial scale (Beever et al. 2006). Alternatively, several recent analyses explicitly create models with hierarchical structure, e.g., with different predictors separated into submodels, so that relationships at disparate scales can be modeled and perhaps combined (Mackey & Lindenmayer 2001). Some Bayesian approaches allow explicit hierarchies and can include process-related elements that might operate across scales (Latimer et al. 2006). Alternatively, hierarchical regression models ("mixed models") allow nested structures of data (Beever et al. 2006), and hierarchical canonical variance partitioning can be used to provide a structured decomposition of variance across scales (Cushman & McGarigal 2002). Unfortunately, the relative merits of these different approaches appear untested both theoretically and practically, and it remains unclear whether more complex hierarchical approaches achieve as much or more than a well-constructed set of predictors used in a sensibly fitted nonhierarchical model. There is ample opportunity to progress knowledge on this topic, particularly with a coherent treatment of theory, data requirements, and model structure.

Supplemental Material

The Interplay of Geographic and Environmental Space

One important concept central to SDMs is the distinction between geographic and environmental space. Whereas geographic space is defined by two-dimensional map coordinates or three-dimensional digital elevation models, environmental space is potentially multi-dimensional, defined by some set of environmental predictors (**Figure 1**). When an SDM is fitted using solely environmental predictors it models variation in occurrence or abundance of a species in environmental space. Any calculation of predictions for new sites is also based on the species' locations in environmental rather than geographic space. Importantly, such a model is effectively ignorant of geographic proximity even when predictions are mapped into geographic space. Mapped

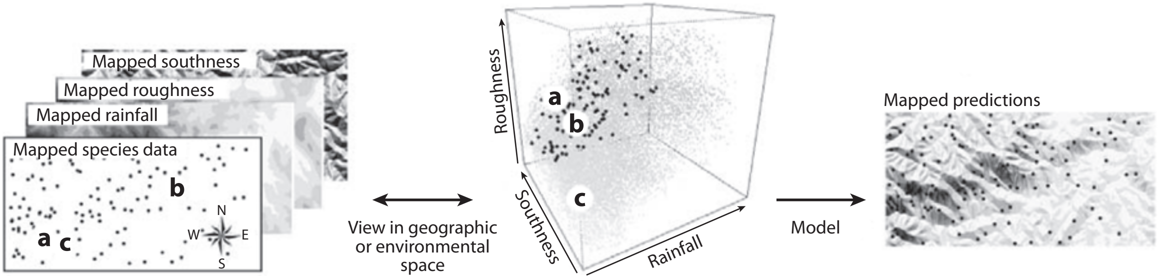

Figure 1

The relationship between mapped species and environmental data (*left*), environmental space (*center*), and mapped predictions from a model only using environmental predictors (*right*). Note that inter-site distances in geographic space might be quite different from those in environmental space—*a* and *c* are close geographically, but not environmentally. The patterning in the predictions reflects the spatial autocorrelation of the environmental predictors.

predictions show clustering and appear spatially informed, but in SDMs with solely environmental predictors this simply reflects the spatial autocorrelation of environment (**Figure 1**).

We note, as an aside, that some SDMs are purely geographic. Examples include geographic range maps, convex hulls, kernel density estimators, kriging, and models of species richness in geographic space. Their use sometimes indicates a belief that geographic processes are dominant over environmental ones, or reflects extremely limited availability of environmental predictors or species data. At most scales and for most species, however, evidence points to the importance of environment in structuring distributions, meaning that inclusion of environment in SDMs is important.

Spatial autocorrelation is an important aspect of the interplay between environmental and geographic space. Geographic clumping of species can result from their response to spatially autocorrelated environmental factors and/or the effects of factors operating primarily in geographic space (Legendre 1993). Where the distribution of a species is largely determined by environmental factors, a properly specified model fitted using an adequate set of predictors will display minimal spatial autocorrelation in its residuals.

Strong residual geographic patterning generally indicates that either key environmental predictors are missing (Leathwick & Whitehead 2001), the model is mis-specified (e.g., only linear terms where nonlinear are required), or geographic factors are influential (Dormann et al. 2007, Miller et al. 2007). The latter include glaciation, fire, contagious disease, connectivity, movement, dispersal, or biotic interactions. For these, the model might require additional relevant predictors, geographic variables and/or realistic estimates of dispersal distances or movement (Ferrier et al. 2002; see **Supplemental Literature Cited**). Alternatively, some modelers enhance SDMs with process-based information to jointly characterize the environmental and spatial influences on distribution (e.g., Rouget & Richardson 2003, Schurr et al. 2007; and see below). Geographic influences in aquatic environments are particularly challenging to model: marine currents can directionally impede dispersal, and in river networks dispersal is generally restricted to the river network and effective distances are strongly influenced by flow directions.

Testing for spatial patterns both in the raw data and model residuals should be part of any SDM study. Methods include use of Moran's I or Geary's c to measure the amount of spatial autocorrelation, addition of local proximity variables to an environmental model to test for residual spatial structure, or use of LISA (local indicator of spatial autocorrelation) to estimate the contribution of each sampling unit to the overall measure of spatial autocorrelation (Dormann et al. 2007, Miller et al. 2007, Rangel et al. 2006).

Alternatively, some approaches explicitly model spatial autocorrelation effects within the modeling process (Rangel et al. 2006). Overall these are used relatively infrequently, although they receive some emphasis in macroecology. One technique is to fit a surface characterizing the geographic pattern (e.g., a trend surface), which is then used as a predictor in the model, sometimes with other environmental predictors used to model the remaining variation (Rangel et al. 2006). Although this describes and controls for geographic pattern it is not fully integrated into the modeling process, and it introduces the risk of confusing geographic effects with spatially autocorrelated environmental terms. More integrated and coherent methods are reviewed in papers detailed in the **Supplemental Literature Cited**; these include autoregressive methods, geostatistical methods based around kriging, generalized linear mixed models, generalized estimating equations, and geographically weighted regression. Currently these methods are more difficult to implement than standard techniques so they are under-utilized, but they have appealing properties and further development might promote their wider use.

Supplemental Material

None of the methods reviewed here provide a strong basis for distinguishing between spatial and environmental effects, though a careful interpretation of the model and its predictions might provide useful insights. Erroneous use of geographic terms to correct for either missing environmental predictors or wrongly specified models is likely to result in poor predictive ability, especially when extrapolating to new regions or times (Dormann et al. 2007, and see below).

Using Models for Explanation versus Prediction

Trends in SDM usage reveal subtle but important shifts in intention. Many early studies had a strong ecological focus, seeking insight, even if indirectly, into the causal drivers of species distributions (Mac Nally 2000). SDMs are still regularly used for such purposes, particularly in quantitative ecological studies (Leathwick & Austin 2001) and evolutionary biology (Graham et al. 2004b). With growing sophistication of modeling algorithms, greater availability of spatially extensive environmental data, and strong demand for mapped products for conservation and land management, an increasing number of papers now focus on predicting distributions (e.g., Hamazaki 2002, and **Supplemental Literature Cited**). Ecological understanding is, of course, still critical to such applications, particularly in the selection of predictors and models and the interpretation of results.

Training data: those data (species records and predictors) used to fit the model

Prediction is used in two main ways. First, predictions are made to new sites within the range of environments sampled by the training data and within the same general time frame as that in which the sampling occurred. We call this model-based interpolation to unsampled sites. Typical applications include global analyses of species distributions, mapping within a region for conservation planning or resource management, and identifying suitable habitat for rare species (Guisan & Thuiller 2005). Such interpolation is usually reliable enough for effective decision making provided that the data and model are reasonable, and any correlations between predictor variables are stable across the geographical domain for which predictions are made.

Second, models are also used to predict to new and unsampled geographic domains and/or future or past climates. The environments in these new times and places need to be carefully assessed, particularly for new combinations of predictor values or for predictor values outside their original ranges in the training data. Prediction to new geographic regions is a special case and has been termed transferability, but often without clear information on the environmental similarities and differences between the model fitting and prediction regions (see **Supplemental Literature Cited**). Prediction to new environments is generally termed extrapolation or forecasting (Araújo & New 2007, Miller et al. 2004). It is inherently risky because no observations of species occurrence are available from the training data to directly support the predictions (see sidebar, Using Models

USING MODELS FOR EXTRAPOLATION

Key assumptions of SDMs are that species are at equilibrium with their environments, and that relevant environmental gradients have been adequately sampled. Use in non-equilibrium settings (e.g., invasions, climate change) usually involves species records unrepresentative of new conditions, and prediction to novel environments. Critics have identified several problems with SDMs and extrapolation, including: different (combinations of) environmental factors may limit distributions or biotic interactions may change substantially in the new context; outcomes will be influenced by genetic variability, phenotypic plasticity and evolutionary changes; dispersal pathways are difficult to predict (De Marco et al. 2008, Dormann 2007, Midgley et al. 2006). However, correlative models currently remain one of few practical approaches for forecasting or hindcasting distributions. We expect that SDMs have a contribution, providing methods and results are rigorously assessed.

Several approaches can improve the use of models for extrapolation, and reduce or expose errors. Differences between the sampled and prediction spaces can be quantified (e.g., similarity measures, Williams et al. 2007; **Figure 2**); species data can be weighted to represent the invasion process or the sample bias of records (Phillips et al. 2009); dispersal can be incorporated using estimates of dispersal rates (Midgley et al. 2006), models of dispersal (Schurr et al. 2007), or by linking SDMs to cellular automata (Iverson et al. 2009); evolutionary change might be estimable and included in models (Hoffmann & Kellermann 2006). Predictions can be tested through retrospectives (Araújo et al. 2005). Differences between models can be reduced by consensus (Pearson et al. 2006), used for discovering why predictions differ (Elith & Graham 2009), or quantified to inform risk analyses and decision making. Alternatively, SDMs can be linked with landscape, population, and physiological models representing processes of change (Kearney & Porter 2009, Keith et al. 2008). Substantial challenges remain, especially those related to how biotic interactions are likely to change and how they can be modeled.

for Extrapolation). As an aside, it is worth recognizing that some researchers exclude interpolation from their definition of prediction, reserving prediction for extrapolation to new conditions or solely for inference from causal models (Berteaux et al. 2006).

A focus on prediction rather than explanation has implications for the way that models are fitted and evaluated. Models for prediction need to balance specific fit to the training data against the generality that enables reliable prediction to new cases. Information criteria such as AIC (Akaike's Information Criterion) address this balance by trading off explained variation against model complexity. Alternatively, data mining and machine learning methods use cross-validation or related methods to test model performance on held out data, both within the model-fitting process, and for model evaluation (Hastie et al. 2009). We anticipate expanding interest in machine learning methods for prediction. The special case of extrapolation needs more attention, so that robust model fitting and testing methods can be developed.

The Need for Functionally Relevant Predictors

Some SDM studies include many candidate predictors, motivated by their ready availability and a belief that the model will identify those that are important. By contrast, a number of modelers have argued strongly for use only of predictors that are ecologically relevant to the target species. Mac Nally (2000) comments: "Statistical tinkering, which really (is) what the entire domain of model selection is about, can never be a substitute for intelligent prior selection of independent variables that may influence the dependent variable.... The variable-selection process will be substantially improved—and, therefore, the inferences too—if that process involves building upon existing knowledge and theory."

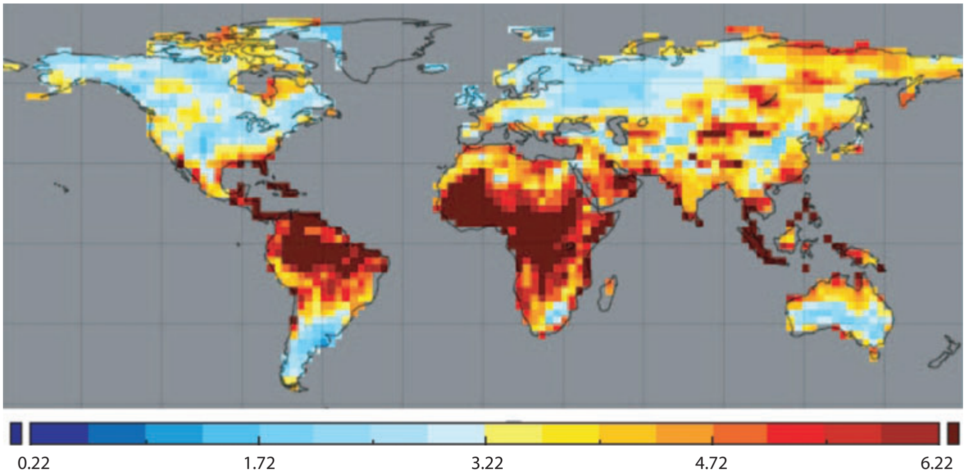

Figure 2

Dissimilarities between 2000 A.D. climates and those (within 500 km of a target site) estimated for 2100 A.D. using multimodel ensembles for the A2 scenario of the IPCC fourth assessment report. High dissimilarities (*red*) indicate the risk of regionally novel climates (from Williams et al. 2007, used with permission).

Austin and Smith (1989, cited in Austin 2002) provide an early example of a deliberate and rigorous approach to predictor selection, distinguishing between resource, direct and indirect gradients. Indirect gradients in terrestrial studies are represented by (distal) predictors such as elevation, which rarely directly affect species distributions. Instead, they are correlated, and sometimes only loosely, with more functionally relevant (proximal) predictors such as temperature, rainfall and solar radiation. In marine systems depth is an indirect proxy for several proximal predictors: temperature and its variability, salinity, light, pressure, and the availability of elements (e.g., calcium).

Use of more ecologically relevant predictors is increasingly possible as interpolated estimates of climate factors and remotely sensed data are more readily available. Franklin (2009, Chapter 5) reviews these predictors comprehensively. Terrestrial examples include Box's analysis of global plant distributions (Box 1981), Zimmermann & Kienast's (1999) use of growing degree days for modeling Swiss tree distribution, and several studies using water balance models of varying sophistication to estimate water availability (see Austin 2007 for a review). Leathwick et al. (2008) constructed functionally relevant predictors of freshwater fish distributions, including estimates of catchment-driven variability in local flow, and access to and from the sea for migratory species. Maravelias & Reid (1997) used surface and seafloor temperature, salinity, and zooplankton availability to predict herring abundance. Remote sensing also offers data that can be adapted to represent proximal predictors—for instance, for approximating habitat complexity for birds (Vierling et al. 2008; St-Louis et al. 2009). Despite these advances, many studies appear to use only data that are readily at hand, failing to explain the relevance of selected predictors, and likely missing important ecological drivers.

While it is logical that ecologically relevant predictors are necessary for explanation and insight, it could be argued that any predictors will suffice if prediction is the sole aim. Multiple lines of evidence suggest otherwise. Predictions show patterned residuals when variables are inadequate, and can be improved substantially by using more proximal predictors (Leathwick & Whitehead 2001), and small data sets and model selection difficulties mean that models can select irrelevant variables (Mac Nally 2000, Steyerberg et al. 1999). Extrapolation in space or time will be particularly error-prone if only distal predictors are used, because the correlations between these and the proximal drivers vary both in space and time (Austin 2002).

Methods for Modeling: Mathematical Form and Fitting Procedures

Supplemental Material

Many methods are used to fit SDMs (Franklin 2009). Although those chosen for particular studies often reflect the nature of the data and/or the question being addressed, some differences between disciplines appear to be driven by "accepted usage," for example, the continued use of GLMs in marine studies and the common use of artificial neural nets (ANNs) for freshwater fish. Historically, the methods used to analyze data sets gathered with intention and design have tended to differ from those using collated records of species records (presence-only data compiled largely opportunistically), but methods are now increasingly convergent. Here we present only a few main points related to analytical approaches; see the **Supplemental Literature Cited** for further reading.

Techniques for modeling very sparse data include convex or alpha hulls (Burgman & Fox 2003), and—where expert opinion is considered more reliable than species records—maps drawn by hand, GIS overlays (combinations of mapped data), or habitat suitability indices (HSIs) (Elith & Leathwick 2009, Franklin 2009).

Some of the earliest numerical SDMs used environmental envelope models to describe the species' range in relation to a set of predictors (Box 1981, Nix 1986). These define the hyper-rectangle that bounds species records in multi-dimensional environmental space, weighting each predictor equally. Such models can be combined with spatially comprehensive environmental data to map likely occurrences, and methods exist for dealing with outliers, e.g., by quantifying percentiles of the distribution. Related techniques (detailed in Franklin 2009) use distance metrics such as the Gower metric or Mahalanobis distance to predict the environmental similarity between records of occurrence and unvisited sites.

Regression-based models extend envelope and similarity approaches by modeling variation in species occurrence or abundance within the occupied environmental space, and selecting predictors according to their observed importance. GLMs were commonly used in early analyses of presence-absence and count data, often with simple additive combinations of linear terms. As the common occurrence of nonlinear species' responses to environment was recognized (Austin et al. 1990), more studies included quadratic, cubic, or other parametric transforms. Generalized additive models (GAMs) are similar to GLMs but use data-defined, scatter plot smoothers to describe nonlinear responses. They have provided useful additional flexibility for fitting ecologically realistic relationships in SDMs.

Regression methods are widely used by ecologists; they can be extended to model complex data types including abundance data with many zeros, records with imperfect detection of presence, and structured samples of data such as sites nested within forest fragments (see **Supplemental Literature Cited**). More generally, many SDM methods are regression-like, assuming that a species' occurrence or abundance can be modeled using additive combinations of predictors, and sometimes also including manually selected terms representing interactions between predictors. Bayesian alternatives are also available (Latimer et al. 2006), bringing sophisticated model-fitting abilities that can incorporate process-based information (e.g., rates of spread; Hooten et al. 2007). However these can require specialized mathematics and programming, and this currently hinders wider uptake despite apparent advantages.

As SDM applications focused more on prediction, researchers looked to methods developed especially for prediction, including those in the machine learning and data mining communities. Examples include ANNs (Olden et al. 2008), multivariate adaptive regression splines (Moisen

& Frescino 2002), classification and regression trees and ensembles of trees (random forests: Prasad et al. 2006; boosted regression trees: Elith et al. 2008), genetic algorithms (Stockwell & Peters 1999), support vector machines (Drake et al. 2006), and maximum entropy models (Phillips et al. 2006). Some of these provide well-controlled variable selection and coefficient estimation, and several are capable of automatically detecting and fitting interactions between predictors. As a consequence their predictive performance may exceed that of more conventional techniques (Elith et al. 2006). While the complex and sometimes "black-box" nature of these techniques has perhaps limited their use, particularly for studies focusing on ecological insight, tools for visualizing and summarizing these models in ways relevant to ecologists are increasingly available. The other immediate constraint to uptake of machine learning techniques is that they are rarely taught in ecological courses, but we expect that to change rapidly in coming years.

Phylogeography: the spatial arrangements of genetic lineages, especially within and among closely related species

Modeling into the Past or the Future

SDMs always have some degree of temporal dimension or reference reflecting their use of species and environmental data gathered over particular time periods (Schröder & Seppelt 2006). However, whereas traditional applications of SDM generally assume a constant and current time frame (even if integrated over some months or years), numerous studies now include temporal change. These target questions relating to recent changes in distributions from disturbances including fire and land use change, the spatial and environmental correlates of speciation events, hybrid zones, paleo-distributions and phylogeography, and forecasts of invasions and distributions under climate change. A key distinction is between those applications requiring predictions in a time period matching that of the training data, compared with those using a model of the current distribution of a species to either hindcast or forecast distributions at some other point in time.

SDMs can explicitly include time as a predictor in the model. For instance, the **Supplemental Literature Cited** lists examples using time-varying food resources in an RSF for grizzly bears, and estimates of time since disturbance for modeling pioneer species in a fire-prone landscape. Models also use retrospective data, e.g., combining historical survey and remnant vegetation records to model pre-clearing vegetation distribution, or modeling pollen records with paleoclimatic data.

SDMs with an evolutionary focus evaluate spatial patterns of inter- and intra-specific variation (see Kozak et al. 2008, Richards et al. 2007, Swenson 2008 for reviews). For instance, the **Supplemental Literature Cited** presents examples that use phylogenetic data and climate envelopes to explore speciation mechanisms in frogs, assess the role of climate in maintaining the location of hybrid zones in birds, and explore species delimitation in salamanders.

Those applications using models to make predictions for time frames substantially different from those of the training data generally require extrapolation in environmental space (see sidebar, Using Models for Extrapolation). Models of the biotic repercussions of global warming and land-use changes require forecasting (Araújo & New 2007, Fitzpatrick et al. 2007, Thuiller et al. 2005), and hindcasting is used for exploring the effects of climate on evolutionary patterns (Kitchener & Dugmore 2000, Kozak et al. 2008, Ruegg et al. 2006). However, understanding and assessing the uncertainties inherent in model predictions for these applications is particularly problematic.

THE SPECIES MODELING LANDSCAPE: HOTSPOTS, RARITIES, AND DIRECTIONS OF CHANGE

Here we identify areas undergoing either rapid development or receiving particularly strong interest, and also explore some less commonly researched topics.

Linking Ecological Theory and Distribution Modeling

Although good linkage between model assumptions and underlying theories and concepts might be reasonably expected in any scientific discipline, several researchers have criticized the SDM community for its lack of theoretical grounding (e.g., Austin 2002, Jiménez-Valverde et al. 2008). In a penetrating critique Huston (2002) states, "[C]ontinued development of rigorous statistical approaches to analyzing habitat data, assisted by the spread of easy computation . . . has been unaccompanied . . . by corresponding development of rigorous logic." Consequences include poorly informed use of models, slow improvement in the ecological realism of methods, and limited uptake of SDM methods and results by other disciplines in which they could be relevant. For instance, recent commentaries by macroecologists and biogeographers (Gaston et al. 2008, Sagarin et al. 2006) point to many interesting theoretical questions about species ranges, but barely refer to insights from the SDM literature, possibly because SDM practitioners largely fail to explicitly identify the broader relevance of their work.

One exception to this general neglect of theoretical issues is a recent debate on the relationship between SDMs and the species niche (see sidebar, The Name Niche Modeling). Unfortunately, this discussion has been plagued by semantic, conceptual, and technical difficulties, and has yet to reach consensus. In common with Austin (2002), Huston (2002) and others, we believe that a more wide-ranging approach to linking theory, data, and models would bring substantial benefits. Important issues additional to niche concepts include the degree of equilibrium in species distributions; how to identify, construct and test functionally relevant predictors; whether current, predominantly

THE NAME NICHE MODELING

Early efforts to relate SDMs to the niche concept were cautious, acknowledging limitations in both data and models. For instance, to Booth et al. (1988), natural distribution data described only the "realized niche," i.e., the competition-mediated distribution. Similarly, Austin et al. (1990) and Austin (2002) described their probabilistic models of eucalypt distribution as an approximation to the "qualitative environmental realized niche," perhaps with sink habitats also included.

Peterson and Soberon have argued for conceptual distinctions between ecological niche models (ENMs) and SDMs, restricting "SDM" to those models containing biotic or accessibility predictors and/or being limited in spatial extent (Peterson 2006). Whilst the links between their framework (Soberon 2007), data types, and models are not yet entirely clear, it appears that they include all environment-based models in their definition of ENMs, particularly (though it's not clear whether exclusively) if absence data are not used. They imply that ENMs get closer to modeling the fundamental niche, but we find this interpretation problematic. In particular, they fail to explain how the methods they class as ENMs technically overcome the well recognized difficulty in describing the fundamental niche from landscape observations of species occurrence.

Other attempts to define what is being modeled have not achieved consensus, partly because definitions of niches are not consistent, and data, methods, and scales overwhelmingly variable (Soberon 2007, Franklin 2009). Araújo & Guisan (2006) question whether the distinction between fundamental and realized niche is useful for these models, given ambiguities in the original formulation of the niche concept. In our view, a more realistic stance is to retain a healthy skepticism about which components of the niche are represented by predictions from an SDM. This is more likely to promote careful analysis of the adequacy of the data used for modeling, while also allowing for uncertainties in predictions and providing impetus for refining understanding through collecting better data, conducting ecological experiments, and testing new ways to model dispersal limitations, effects of competitors, and so on. Use of neutral terminology to describe species distributional models (SDM rather than ENM) seems preferable.

additive, modeling methods are ecologically realistic (see, for instance, multiplicative models: McCune 2006); how to deal with interspecific interactions; and how to understand and model the interplay of geographic and environmental drivers of species distributions across different spatial and temporal scales.

Pseudoabsence: a location at which predictors are sampled, variably viewed as a sample of the "background" or sampling universe, or an implied absence

When Absence Is Not Known

Presence-only data consist of records describing known occurrences (presence) of species, but lacking information about known absences. One example is the radiotelemetry data collected in wildlife studies. Analysis of these data with use-availability models has received steady attention over recent years (Pearce & Boyce 2006). Alternatively, museum records are now often utilized for evolutionary biology, macroecology, conservation, invasive species, and climate change modeling, using the millions of records compiled in electronic form from natural history collections (Graham et al. 2004a). Despite their limitations, use of such data is often justified by the lack of systematic survey data, coupled with widespread demand for mapped predictions.

Modelers are still coming to terms with how best to model presence-only data. Where analytical methods were once restricted to envelopes and distance measures, comparison of presence records with background or pseudoabsence points is now common (e.g., using GARP, ENFA, MaxEnt, and regression methods). Reviews and comparisons include Franklin (2009) and Elith et al. (2006). Attitudes to the value of presence-only data are remarkably variable. Some acknowledge that their predictions would be more robust if presence-absence or abundance data were available—a view that, if accepted, has substantial implications for the type of data that ecologists should aim to collect. An advantage of presence-absence data is that it conveys valuable information about surveyed locations (enabling analyses of biases) and prevalence (Phillips et al. 2009). Others argue that absence records introduce confounding information because they can indicate either habitat that is unsuitable or habitat that is suitable but is unoccupied, perhaps because of inaccessibility. This idea is commonly linked to the concept of modeling potential distributions (Jiménez-Valverde et al. 2008). Absence data are also sometimes viewed as misleading because the species or environment is not at equilibrium (e.g., invasions, climate change) or the species not easily detected. Interpretation of the meaning of background data or pseudoabsence data also varies. In general, the literature lacks robust discussion of the interplay between these disparate views and ecological and statistical theory. Progress in these topics, and on methods for detecting and dealing with sample bias and for evaluating presence-only models, could bring substantial benefits.

Modeling Responses Other than the Mean

Most methods for modeling presence-absence or abundance data estimate the center of the conditional distribution of the response, or the mean. Some argue that a more complete summary of the quantiles of the conditional distribution is useful (Austin 2007, Huston 2002). Upper quantiles, those near the maximum response, have received the most attention, based on the assumption that they better represent the response of the species to a predictor when other variables are not limiting (Huston 2002). They can reveal biases or missing predictors, and arguably can indicate the potential rather than the actual distribution (Cade et al. 2005). Low quantiles might also be relevant—for example, to estimate the lowest recruitment level for a species (Planque & Buffaz 2008). Interesting recent applications (see **Supplemental Literature Cited**) include freshwater, marine, and phylogenetic studies. So far, ecological examples are limited to parametric or nonparametric regression and gaussian responses, but methods are emerging that use tree ensembles and k-nearest neighbors and/or allow for differing response types (see **Supplemental Literature Cited**).

Supplemental Material

Biotic Interactions

Very few SDM studies explicitly include predictors describing biological interactions (Guisan & Thuiller 2005). In one early study, Austin & Cunningham (1981) included terms describing the presence of conspecifics in models of eucalypts, whilst acknowledging the possibility that variation attributed to conspecifics might reflect some missing but unknown environmental predictors. This typifies the difficulty in making inferences about the relative importance of jointly fitted abiotic and biotic predictors (Guisan & Thuiller 2005), because in most data sets environmental effects are confounded with those of competitors and mutualists. One exception is provided by Leathwick & Austin (2001) who treated geographic disjunctions in New Zealand's *Nothofagus* forests as a "natural removal experiment." Their SDMs indicated high levels of competitive interaction, with this effect varying depending on environmental conditions.

Given these difficulties, most practitioners use abiotic predictors alone. In models for understanding or interpolation-style prediction, the consequences may not be too severe, except where the presence of a host species is critical (e.g., Wharton & Kriticos 2004) and not predicted by the available covariates. However, for extrapolation (e.g., global warming, invasions), the effects of competitors, mutualists, and conspecific attractions might have far-reaching effects, especially where novel combinations of species are likely to occur (see sidebar, Using Models for Extrapolation). This is one of the more difficult aspects of SDMs, and we anticipate that its resolution will most likely require development of methods with capabilities beyond those available in current methods.

Integrating Pattern and Process

Several groups are now exploring how to better represent ecological processes within correlative models (see Schröder & Seppelt 2006 for a review), particularly for nonequilibrial situations. For example, Rouget & Richardson (2003) modeled the abundance of an invader allowing effects of propagule pressure; Hooten et al. (2007) modeled spread of the Eurasian collared dove using a hierachical Bayesian model incorporating density-dependent growth and dispersal, and Iverson et al. (2009) modeled emerald ash borer movement within predicted distributional ranges of trees. Others suggest combining SDMs with different types of models that allow inclusion of mechanistic, population, and landscape change effects (Drielsma & Ferrier 2009, Kearney et al. 2008, Keith et al. 2008).

Model Selection

Early SDMs generally used statistical techniques based on p-values for model selection, but a recent shift has seen much greater emphasis on AIC and multimodel inference (Burnham & Anderson 2002). This shift has been useful for reducing reliance on the "truth" of a model selected by stepwise procedures and for understanding the error tendencies of conventional selection approaches (Whittingham et al. 2006). However, though this type of multimodel inference is useful for exploring model-based uncertainty, whether it is the best way to reliably predict an outcome is unclear. Other model averaging techniques from computer science use a range of approaches to concurrently develop a set of models that together predict well (Hastie et al. 2009). Research comparing the conceptual bases and performance of various model averaging approaches including regression/AIC, Bayesian methods, and machine learning model ensembles (e.g., bagged or boosted trees, Prasad et al. 2006) could be profitable.

There are also interesting alternative approaches to selecting a single final model. The different information criteria provide a range of trade-offs between model complexity and predictive performance and can be used within cross-validation to select a model (Hastie et al. 2009). Some methods focus on simultaneous selection of variables and parameter estimation, for example, by shrinking coefficient estimates (e.g., see Reineking & Schröder 2006 on ridge regression and the lasso). These provide alternative methods for selecting a final regression model that are generally more reliable than stepwise methods. In machine learning these ideas of model selection and tuning are termed "regularization," i.e., making the fitted surface more regular or smooth by controlling overfitting (e.g., used in MaxEnt, Phillips et al. 2006). Use of these alternative model selection methods in ecology are still relatively rare, but likely to increase.

Model Evaluation

Supplemental Material

Although the need for robust model evaluation is widely acknowledged, there are diverse opinions on what properties of a model are important and how to test them appropriately (see **Supplemental Literature Cited**). Where modelers aim to explain patterns or generate hypotheses (e.g., in evolutionary biology and classical ecological studies), results are generally assessed using statistical tests of model fit and comparison with existing knowledge. In contrast, when prediction is the aim, evaluation targets predictive ability and current practice usually involves testing predictive performance using data resampling (split samples, cross-validation, bootstrapping) or, more rarely, independent data sets. Most summaries of performance are based on a relatively small set of statistics including kappa, area under the receiver operating characteristic curve (AUC) and correlation coefficients. Several researchers have attempted to understand the relative performance of these tests including their sensitivity to data characteristics, but progress toward adoption of a comprehensive toolbox of evaluation measures is slow and impeded by arguments about the general validity of some statistics. Instead, it would be more constructive to identify the proper place of each statistic in the broad realm of what needs testing. The machine learning and weather-forecasting communities have developed expertise in testing predictive performance and use some statistics rarely considered in ecology (Caruana & Niculescu-Mizil 2006, Pearce & Ferrier 2000; see also **Supplemental Literature Cited**). SDM evaluation would benefit from identifying useful techniques in other fields, and from more research focus on topics such as how to analyze spatial patterns in errors, how to deal with uncertainties, and how to assess model performance in the context of the intended application, including decision making. More use of artificial data (Austin 2007) and more experimental verification of modeled relationships (e.g., Wright et al. 2006) could also yield valuable insights.

Uncertainty

Use of SDM for applications such as conservation planning and biosecurity creates an imperative for considering errors and their relative costs. Uncertainty in SDMs results both from data deficiencies (e.g., missing covariates, and samples of species occurrences that are small, biased, or lacking absences) and from errors in specification of the model (Barry & Elith 2006). A few papers provide taxonomies of uncertainty as a basis for assessing errors, and suggest general treatments. Heikkinen et al. (2006) review various aspects of SDMs that contribute to uncertainty; Hortal et al. (2008) provide a commentary on biodiversity data and its uncertainties; and Burgman et al. (2005) review treatments of uncertainty in landscape ecology. Relatively few studies address uncertainty in SDMs and its effects on the model, predictions, and related decision making (but see **Supplemental Literature Cited**). Model uncertainty has received most attention, particularly in the context of model averaging or consensus, but also for providing mapped uncertainty estimates. Studies on data errors include assessments of the influence of errors and biases in species records, and in predictors such as digital elevation models and their products. These extend beyond the uncertainty that can be estimated from standard errors of parameters in a regression model, or from bootstrapped estimates of uncertainty. Modelers can attempt to reduce uncertainty, and/or characterize it and explore its effects on decision making. Because problems related to uncertainty are difficult to deal with they are often ignored, but we anticipate increasing recognition of their importance, particularly in management applications.

CONCLUDING THOUGHTS

Reflection on the broad scope of both past and current SDM writings reveals a rich diversity of biological and environmental settings, philosophical and analytical approaches, and research and management applications. Our summary of this multifaceted and developing field may disappoint those looking for specific advice or a more methodologically oriented review—we regard a number of emerging books and teaching resources as better able to fill these needs. Our emphasis reflects the belief that further advances in SDM are more likely to come from better integration of theory, concepts, and practice than from improved methods per se. Our hope is that this review will encourage more deliberate exploration across discipline boundaries, the informed and creative use of a breadth of approaches, and planned endeavors to fill important knowledge gaps. This expanded focus should, in turn, improve the ability of SDMs to make their contribution to delivering the type of information required for managing the Earth's dwindling biological resources.

SUMMARY POINTS

1. Modern SDMs represent the convergence of site-based ecology and advances in GIS and spatial data technologies. They are applied across terrestrial, freshwater, and marine environments, at widely varying spatial and temporal scales, and to gain ecological and evolutionary insight and predict distributions. Differences in mobility between species motivate some of the most marked differences in modeling approach.
2. Species distributions reflect the interplay of geographic and environmental processes. Using ecologically relevant environmental variables and addressing residual geographic patterning are both important.
3. Prediction takes two forms: interpolation and extrapolation. The latter violates several statistical and ecological assumptions of SDMs, so hindcasting (evolutionary questions) and forecasting (climate change and invasive species models) require special care.
4. Development of stronger links between ecological theory and concepts and SDM practice would be beneficial for developing more robust and consistent use of these techniques.

FUTURE ISSUES

1. Methods are required for dealing with uncertainty: characterizing it, reducing it, or assessing its influence on decisions.
2. Model selection and evaluation methods are likely to expand and incorporate new techniques from statistics, weather forecasting, and machine learning.

3. The use of presence-only data will continue, so methods for dealing with biases and evaluating results need more development.
4. Cycles of development, implementation, and evaluation (including experimental testing) would provide insights, strengthen links to theory, and contribute important information for developing ecologically relevant predictors.
5. Many applications could benefit from advances in modeling biotic interactions and other ecological processes.
6. If SDMs are to be used for extrapolation, more assessments of whether they are fit for purpose are required. We need carefully targeted studies addressing performance across different spatial and temporal scales and degrees of equilibrium, in the context of the nature of actions that will flow from the predictions.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The preparation of this review was greatly assisted by generous and perceptive readers in Mike Austin, Mark Burgman, Yung En Chee, Simon Ferrier, Janet Franklin, Catherine Graham, and Andrea White. We have also benefitted enormously over the years from the insights of colleagues across the world who have challenged us, encouraged us, and helped us to learn. J.L.'s contribution was funded partly by New Zealand's Foundation for Research, Science and Technology under Contract C01×052, and J.E.'s by ARC grant DP0772671 and the Australian Center of Excellence for Risk Analysis.

LITERATURE CITED

Addicott JF, Aho JM, Antolin MF, Padilla DK, Richardson JS, et al. 1987. Ecological neighborhoods: scaling environmental patterns. *Oikos* 49:340–46

Allen TFH, Starr TB. 1982. *Hierarchy: Perspectives for Ecological Complexity*. Chicago: Univ. Chicago Press

Araújo MB, Guisan A. 2006. Five (or so) challenges for species distribution modelling. *J. Biogeogr.* 33:1677–88

Araújo MB, New M. 2007. Ensemble forecasting of species distributions. *Trends Ecol. Evol.* 22:42–47

Araújo MB, Pearson RG, Thuiller W, Erhard M. 2005. Validation of species-climate impact models under climate change. *Global Change Biol.* 11:1504–13

Austin MP. 1985. Continuum concept, ordination methods and niche theory. *Annu. Rev. Ecol. Syst.* 16:39–61

Austin MP. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecol. Model.* 157:101–18

Austin MP. 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecol. Model.* 200:1–19

Austin MP, Cunningham RB. 1981. Observational analysis of environmental gradients. *Proc. Ecol. Soc. Aust.* 11:109–19

Austin MP, Nicholls AO, Margules CR. 1990. Measurement of the realized qualitative niche: environmental niches of five eucalypt species. *Ecol. Monogr.* 60:161–77

Barry SC, Elith J. 2006. Error and uncertainty in habitat models. *J. Appl. Ecol.* 43:413–23

Beever EA, Swihart RK, Bestelmeyer BT. 2006. Linking the concept of scale to studies of biological diversity: evolving approaches and tools. *Divers. Distrib.* 12:229–35

Berteaux D, Humphries MM, Krebs CJ, Lima M, McAdam AG, et al. 2006. Constraints to projecting the effects of climate change on mammals. *Clim. Res.* 32:151–58

Booth TH, Nix HA, Hutchinson MF, Jovanovic T. 1988. Niche analysis and tree species introduction. *For. Ecol. Manag.* 23:47–59

Box EO. 1981. Predicting physiognomic vegetation types with climate variables. *Vegetatio* 45:127–39

Burgman MA, Fox JC. 2003. Bias in species range estimates from minimum convex polygons: implications for conservation and options for improved planning. *Anim. Conserv.* 6:19–28

Burgman MA, Lindenmayer DB, Elith J. 2005. Managing landscapes for conservation under uncertainty. *Ecology* 86:2007–17

Burnham KP, Anderson DR. 2002. *Model Selection and Inference: A Practical Information-Theoretic Approach*. New York: Springer-Verlag. 488 pp. 2nd ed.

Cade BS, Noon BR, Flather CH. 2005. Quantile regression reveals hidden bias and uncertainty in habitat models. *Ecology* 86:786–800

Capen DE, ed. 1981. The use of multivariate statistics in studies of wildlife habitat. *Gen.Tech. Rep. RM-87*, Rocky Mt. For. Range Exp. Stn., USDA For. Serv.

Caruana R, Niculescu-Mizil A. 2006. *An empirical comparison of supervised learning algorithms*. Presented at Proc. Int. Conf. Machine Learn., 23rd, Pittsburgh, PA

Cushman SA, McGarigal K. 2002. Hierarchical, multi-scale decomposition of species-environment relationships. *Landscape Ecol.* 17:637–46

De Marco P, Diniz-Filho JAF, Bini LM. 2008. Spatial analysis improves species distribution modelling during range expansion. *Biol. Lett.* 4:577–80

Dormann CF. 2007. Promising the future? Global change projections of species distributions. *Basic Appl. Ecol.* 8:387–97

Dormann CF, McPherson JM, Araujo MB, Bivand R, Bolliger J, et al. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30:609–28

Drake JM, Randin C, Guisan A. 2006. Modelling ecological niches with support vector machines. *J. Appl. Ecol.* 43:424

Drielsma M, Ferrier S. 2009. Rapid evaluation of metapopulation persistence in highly variegated landscapes. *Biol. Conserv.* 142:529–40

Dungan JL, Perry JN, Dale MRT, Legendre P, Citron-Pousty S, et al. 2002. A balanced view of scale in spatial statistical analysis. *Ecography* 25:626–40

Elith J, Graham C. 2009. Do they? How do they? WHY do they differ? ... on finding reasons for differing performances of species distribution models. *Ecography* 32:66–77

Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129–51

Elith J, Leathwick JR. 2009. Conservation prioritization using species distribution models. In *Spatial Conservation Prioritization: Quantitative Methods and Computational Tools*, ed. A Moilanen, KA Wilson, HP Possingham. Oxford: Oxford Univ. Press. pp 70–93

Elith J, Leathwick JR, Hastie T. 2008. A working guide to boosted regression trees. *J. Anim. Ecol.* 77:802–13

Ferrier S, Guisan A. 2006. Spatial modelling of biodiversity at the community level. *J. Appl. Ecol.* 43:393–404

Ferrier S, Watson G. 1997. An evaluation of the effectiveness of environmental surrogates and modelling techniques in predicting the distribution of biological diversity. *Consult. Rep. NSW Natl. Parks Wildl. Serv. Dep. Environ., Sport Territ., Environ.* Aust., Canberra. **http://www.deh.gov.au/biodiversity/publications/technical/surrogates/**

Ferrier S, Watson G, Pearce J, Drielsma M. 2002. Extended statistical approaches to modelling spatial pattern in biodiversity: the north-east New South Wales experience. I. Species-level modelling. *Biodivers. Conserv.* 11:2275–307

Fitzpatrick MC, Weltzin JF, Sanders NJ, Dunn RR. 2007. The biogeography of prediction error: Why does the introduced range of the fire ant overpredict its native range? *Global Ecol. Biogeog.* 16:24–33

Fleishman E, MacNally R, Fay JP, Murphy DD. 2001. Modeling and predicting species occurrence using broad-scale environmental variables: an example with butterflies of the Great Basin. *Conserv. Biol.* 15:1674–85

Foody GM. 2008. GIS: biodiversity applications. *Prog. Phys. Geog.* 32:223–35

Franklin J. 2009. *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge, UK: Cambridge Univ. Press. In press

Gaston KJ, Chown SL, Evans KL. 2008. Ecogeographical rules: elements of a synthesis. *J. Biogeogr.* 35:483–500

Graham CH, Ferrier S, Huettman F, Moritz C, Peterson AT. 2004a. New developments in museum-based informatics and applications in biodiversity analysis. *Trends Ecol. Evol.* 19:497–503

Graham CH, Ron SR, Santos JC, Schneider CJ, Moritz C. 2004b. Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution* 58:1781–93

Grinnell J. 1904. The origin and distribution of the chestnut-backed chickadee. *Auk* 21:364–65

Guisan A, Thuiller W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* 8:993–1009

Guisan A, Zimmermann NE. 2000. Predictive habitat distribution models in ecology. *Ecol. Model.* 135:147–86

Hamazaki T. 2002. Spatiotemporal prediction models of cetacean habitats in the mid-western North Atlantic ocean (From Cape Hatteras, North Carolina, USA to Nova Scotia, Canada). *Mar. Mamm. Sci.* 18:920–39

Hastie T, Tibshirani R, Friedman JH. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* New York: Springer-Verlag. 744 pp. 2nd ed.

Heikkinen R, Luoto M, Araújo MB, Virkkala R, Thuiller W, Sykes MT. 2006. Methods and uncertainties in bioclimatic envelope modelling under climate change. *Prog. Phys. Geog.* 30:751–77

Hoffmann A, Kellermann V. 2006. Revisiting heritable variation and limits to species distribution: recent developments. *Isr. J. Ecol. Evol.* 52:247–61

Hooten MB, Wikle CK, Dorazio RM, Royle JA. 2007. Hierarchical spatiotemporal matrix models for characterizing invasions. *Biometrics* 63:558–67

Hortal J, Jiménez-Valverde A, Gómez JF, Lobo JM, Baselga A. 2008. Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos* 117:847–58

Huston MA. 2002. Critical issues for improving predictions. In *Predicting Species Occurrences: Issues of Accuracy and Scale*, ed. JM Scott, PJ Heglund, ML Morrison, MG Raphael, WA Wall, et al., pp. 7–24. Covelo, CA: Island Press

Iverson LR, Prasad AM, Bossenbroek J, Sydnor D, Schwartz MW. 2009. Modeling potential movements of an ash threat: the emerald ash borer. In *Advances in Threat Assessment and Their Application to Forest and Rangeland Management*. **http://www.threats.forestencyclopedia.net**, ed. J Pye, M Raucher

Jiménez-Valverde A, Lobo JM, Hortal J. 2008. Not as good as they seem: the importance of concepts in species distribution modelling. *Divers. Distrib.* 14:885–90

Jones J. 2001. Habitat selection studies in avian ecology: a critical review. *Auk* 118:557–62

Kearney M, Phillips BL, Tracy CR, Christian KA, Betts G, et al. 2008. Modelling species distributions without using species distributions: the cane toad in Australia under current and future climates. *Ecography* 31:423–34

Kearney M, Porter WP. 2009. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecol. Lett.* 12:334–50

Keith DA, Akçakaya HR, Thuiller W, Midgley GF, Pearson RG, et al. 2008. Predicting extinction risks under climate change: coupling stochastic population models with dynamic bioclimatic habitat models. *Biol. Lett.* 4:560–63

Kitchener AC, Dugmore AJ. 2000. Biogeographical change in the tiger, *Panthera tigris*. *Anim. Conserv.* 3:113–24

Kozak KH, Graham CH, Wiens JJ. 2008. Integrating GIS-based environmental data into evolutionary biology. *Trends Ecol. Evol.* 23:141–48

Latimer AM, Wu SS, Gelfand AE, Silander JA. 2006. Building statistical models to analyze species distributions. *Ecol. Appl.* 16:33–50

Leathwick JR, Austin MP. 2001. Competitive interactions between tree species in New Zealand's old-growth indigenous forests. *Ecology* 82:2560–73

Leathwick JR, Elith J, Chadderton L, Rowe D, Hastie T. 2008. Dispersal, disturbance, and the contrasting biogeographies of New Zealand's diadromous and nondiadromous fish species. *J. Biogeogr.* 35:1481–97

Leathwick JR, Whitehead D. 2001. Soil and atmospheric water deficits and the distribution of New Zealand's indigenous tree species. *Funct. Ecol.* 15:233–42

Legendre P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* 74:1659–73

Levin SA. 1992. The problem of pattern and scale in ecology. *Ecology* 73:1943–67

Mac Nally R. 2000. Regression and model-building in conservation biology, biogeography and ecology: the distinction between—and reconciliation of—'predictive' and 'explanatory' models. *Biodivers. Conserv.* 9:655–71

MacArthur RH. 1958. Population ecology of some warblers of northeastern coniferous forests. *Ecology* 39:599–619

MacKenzie DI, Nichols JD, Lachman GB, Droege S, Royle JA, et al. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83:2248–55

Mackey BG, Lindenmayer DB. 2001. Towards a hierachical framework for modelling the spatial distribution of animals. *J. Biogeogr.* 28:1147–66

Manly BFJ, McDonald LL, Thomas DL, McDonald TL, Erickson WP. 2002. *Resource Selection by Animals: Statistical Design and Analysis for Field Studies.* Dordrecht: Kluwer. 221 pp. 2nd ed.

Maravelias CD, Reid DG. 1997. Identifying the effects of oceanographic features and zooplankton on prespawning herring abundance using generalised additive models. *Mar. Ecol. Prog. Ser.* 147:1–9

McCune B. 2006. Non-parametric models with automatic interactions. *J. Veg. Sci.* 17:819–30

Midgley GF, Hughes GO, Thuiller W, Rebelo AG. 2006. Migration rate limitations on climate change-induced range shifts in Cape Proteaceae. *Divers. Distrib.* 12:555

Miller J, Franklin J, Aspinall R. 2007. Incorporating spatial dependence in predictive vegetation models. *Ecol. Model.* 202:225–42

Miller JR, Turner MG, Smithwick EAH, Dent CL, Stanley EH. 2004. Spatial extrapolation: the science of predicting ecological patterns and processes. *BioScience* 54:310–20

Moisen GG, Frescino TS. 2002. Comparing five modeling techniques for predicting forest characteristics. *Ecol. Model.* 157:209–25

Murray A. 1866. *The Geographical Distribution of Mammals*. London: Day & Son

Nix H. 1986. A biogeographic analysis of Australian elapid snakes. In *Atlas of Elapid Snakes of Australia*, ed. R Longmore, pp. 4–15. Canberra: Aust. Gov. Publ. Serv.

Olden JD, Lawler JJ, Poff NL. 2008. Machine learning methods without tears: a primer for ecologists. *Q. Rev. Biol.* 83:171–93

Pearce J, Ferrier S. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecol. Model.* 133:225–45

Pearce JL, Boyce MS. 2006. Modelling distribution and abundance with presence-only data. *J. Appl. Ecol.* 43:405–12

Pearson RG. 2007. Species' Distribution Modeling for Conservation Educators and Practitioners. Synthesis. New York: Am. Mus. Natl. Hist. **http://ncep.amnh.org**

Pearson RG, Dawson TP. 2003. Predicting the impacts of climate change on the distribution of species: Are bioclimate envelope models useful? *Global Ecol. Biogeog.* 12:361–71

Pearson RG, Thuiller W, Araújo MB, Martinez-Meyer E, Brotons L, et al. 2006. Model-based uncertainty in species range prediction. *J. Biogeogr.* 33:1704–11

Peterson AT. 2006. Uses and requirements of ecological niche models and related distributional models. *Bioinformatics* 3:59–72

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190:231–59

Phillips SJ, Dudik M, Elith J, Graham C, Lehmann A, et al. 2009. Sample selection bias and presence-only models of species distributions. *Ecol. Appl.* 19:181–97

Planque B, Buffaz L. 2008. Quantile regression models for fish recruitment environment relationships: four case studies. *Mar. Ecol. Prog. Ser.* 357:213–23

Poff NL. 1997. Landscape filters and species traits: towards mechanistic understanding and prediction in stream ecology. *J. North Am. Benthol. Soc.* 16:391–409

Prasad AM, Iverson LR, Liaw A. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9:181–99

Rangel TFLVB, Diniz-Filho JAF, Bini LM. 2006. Towards an integrated computational tool for spatial analysis in macroecology and biogeography. *Global Ecol. Biogeog.* 15:321–27

Reineking B, Schröder B. 2006. Constrain to perform: regularization of habitat models. *Ecol. Model.* 193:675–90

Richards CL, Carstens BC, Lacey Knowles L. 2007. Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. *J. Biogeogr.* 34:1833–45

Rouget M, Richardson DM. 2003. Inferring process from pattern in plant invasions: a semimechanistic model incorporating propagule pressure and environmental factors. *Am. Nat.* 162:713–24

Royle JA, Dawson DK, Bates S. 2004. Modeling abundance effects in distance sampling. *Ecology* 85:1591–97

Ruegg KC, Hijmans RJ, Moritz C. 2006. Climate change and the origin of migratory pathways in the Swainson's thrush, *Catharus ustulatus*. *J. Biogeogr.* 33:1172–82

Sagarin RD, Gaines SD, Gaylord B. 2006. Moving beyond assumptions to understand abundance distributions across the ranges of species. *Trends Ecol. Evol.* 21:524–30

Saunders SC, Chen J, Drummer TD, Gustafson EJ, Brosofske KD. 2005. Identifying scales of pattern in ecological data: a comparison of lacunarity, spectral and wavelet analyses. *Ecol. Complex.* 2:87–105

Schimper AFW. 1903. *Plant-Geography upon a Physiological Basis*. Transl. WR Fisher. Oxford: Clarendon Press (From German)

Schröder B. 2008. Challenges of species distribution modeling belowground. *J. Plant Nutr. Soil Sci.* 171:325–37

Schröder B, Seppelt R. 2006. Analysis of pattern–process interactions based on landscape models—overview, general concepts, and methodological issues. *Ecol. Model.* 199:505–16

Schurr FM, Midgley GF, Rebelo AG, Reeves G, Poschlod P, et al. 2007. Colonization and persistence ability explain the extent to which plant species fill their potential range. *Global Ecol. Biogeog.* 16:449–59

Soberon J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. *Ecol. Lett.* 10:1115–23

St-Louis V, Pidgeon AM, Clayton MK, Locke BA, Bash D, et al. 2009. Satellite image texture and a vegetation index predict avian biodiversity in the Chihuahuan Desert of New Mexico. *Ecography* 32:468–80

Stauffer DE. 2002. Linking populations and habitats: Where have we been? Where are we going? In *Predicting Species Occurrences: Issues of Accuracy and Scale*, ed. JM Scott, PJ Heglund, ML Morrison, MG Raphael, WA Wall, et al., pp. 53–61. Covelo, CA: Island Press

Steyerberg EW, Eijkemans MJC, Habbema JDF. 1999. Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. *J. Clin. Epidemiol.* 52:935–42

Stockwell D, Peters D. 1999. The GARP modelling system: problems and solutions to automated spatial prediction. *Int. J. Geogr. Inf. Sci.* 13:143–58

Swenson NG. 2008. The past and future influence of geographic information systems on hybrid zone, phylogeographic and speciation research. *J. Evol. Biol.* 21:421–34

Thuiller W, Richardson DM, Pysek P, Midgley GF, Hughes GO, et al. 2005. Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Global Change Biol.* 11:2234–50

Tobalske C. 2002. Effects of spatial scale on the predictive ability of habitat models for the Green Woodpecker in Switzerland. In *Predicting Species Occurrences: Issues of Accuracy and Scale*, ed. JM Scott, PJ Heglund, ML Morrison, MG Raphael, WA Wall, et al., pp. 197–204. Covelo, CA: Island Press

Venables WN, Dichmont CM. 2004. GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. *Fish. Res.* 70:319–37

Vierling KT, Vierling LA, Gould WA, Martinuzzi S, Clawges RM. 2008. Lidar: shedding new light on habitat characterization and modeling. *Front. Ecol. Environ.* 6:90–98

Wharton TN, Kriticos DJ. 2004. The fundamental and realized niche of the Monterey Pine aphid, *Essigella californica* (Essig) (Hemiptera : Aphididae): implications for managing softwood plantations in Australia. *Divers. Distrib.* 10:253–62

Whittaker RJ. 1956. Vegetation of the Great Smoky Mountains. *Ecol. Monogr.* 26:1–80

Whittingham MJ, Stephens PA, Bradbury RB, Freckleton RP. 2006. Why do we still use stepwise modelling in ecology and behavior? *J. Anim. Ecol.* 75:1182–89

Williams JW, Jackson ST, Kutzbac JE. 2007. Projected distributions of novel and disappearing climates by 2100 AD. *Proc. Natl. Acad. Sci. USA* 104:5738–42

Wright JW, Davies KF, Lau JA, McCall AC, McKay JK. 2006. Experimental verification of ecological niche modeling in a heterogeneous environment. *Ecology* 87:2433–39

Zimmermann NE, Kienast F. 1999. Predictive mapping of Alpine grasslands in Switzerland: species versus community approach. *J. Veg. Sci.* 10:469–82

Annual Review of Ecology, Evolution, and Systematics

Volume 40, 2009

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Ecology, Evolution, and Systematics* articles may be found at http://ecolsys.annualreviews.org/errata.shtml