Cambridge University Press
978-0-521-70002-3 - Mapping species distributions: Spatial inference and prediction
Janet Franklin and Jennifer A. Miller
Excerpt
More information

Part I

History and ecological basis of species distribution modeling

Recent decades have seen an explosion of interest in species distribution modeling. This has resulted from a confluence of the growing need for information on the geographical distribution of biodiversity and new and improved techniques and data suitable for addressing this information need – remote sensing, global positioning system technology, geographic information systems, and statistical learning methods. Developments in this area are occurring so rapidly that it was difficult to know when to stop writing this book. It is very challenging to write about a moving target. For the same reason, however, it was the right time to summarize the foundations of, and recent developments in this enterprise called species distribution modeling (SDM). This book provides an introduction to SDM for beginners in the field and for those wishing to use such models in environmental assessment and biodiversity conservation, while providing a significant reference on current practice for researchers.

In this Part, Chapter 1 establishes some basic terminology used to describe species' distribution modeling, describes a framework for implementing SDM that will be used as an organizing principle for the book, and reviews the problems and applications that have motivated a growing interest in SDM. Much of this book describes a modeling approach that links species location information with environmental data (Part II) in order to quantify the distributions of species on environmental gradients and map those distributions onto geographical space using a model (Part III). Before going into detail about this quantitative approach to describing and modeling species–environment relationships, Chapter 2 reviews other kinds of data and methods that can and have been used to more directly determine species distributions. A growing number of species atlas projects provide wall-to-wall, direct observational data of species distributions at certain scales. Cartographic interpolation and geostatistical techniques can be applied to species location data without reference to environmental data. Chapter 2 discusses the strengths and limitations of using species distribution data derived from these interpolation approaches. Chapter 3 presents the ecological concepts that underlie our understanding of how environmental factors limit species distributions and determine species diversity. Based on this conceptual understanding, suitable data (Part II) and models (Part III) can be chosen and implemented to describe species–environment relationships, and, importantly, the assumptions and limitations of the resulting models and their predictions can be explicitly understood (Part IV).

Cambridge University Press
978-0-521-70002-3 - Mapping species distributions: Spatial inference and prediction
Janet Franklin and Jennifer A. Miller
Excerpt
More information

1 · *Species distribution modeling*

Prediction is very difficult, especially about the future.
Niels Bohr

1.1 Introduction

What are predictive maps of species distributions? Why make them? How? Environmental scientists increasingly need to use local measurements to assess change at landscape, regional and global scales, and statistical or simulation models are often used to extrapolate environmental data in space (Miller *et al.*, 2004; Peters *et al.*, 2004). *Species distribution modeling* (SDM) is just one example of this, but an increasingly important one – SDM extrapolates species distribution data in space and time, usually based on a statistical model. Developing a species distribution model begins with observations of species occurrences, and with environmental variables thought to influence habitat suitability and therefore species distribution. The model can be a quantitative or rule-based model and, if the fit is good between the species distribution and the predictors that are examined, this can provide insight into species environmental tolerances or habitat preferences. It also provides the opportunity to make a spatial prediction. *Predictive mapping*, or geographical extrapolation using the model, results in a spatially explicit "wall-to-wall" prediction of species distribution or habitat suitability (Fig. 1.1). Maps of environmental predictors, or their surrogates, must be available in order for predictive mapping to be implemented (Franklin, 1995).

The purpose of this book is to describe the process of species distribution modeling, review the significant advancements that have been made recently in this field, and try to provide guidelines and a framework for choosing data, methods and study design that are appropriate to a given SDM application. This book provides an introduction to SDM for beginners in the field who wish to use such models in environmental analyses while also providing a comprehensive review of current practice

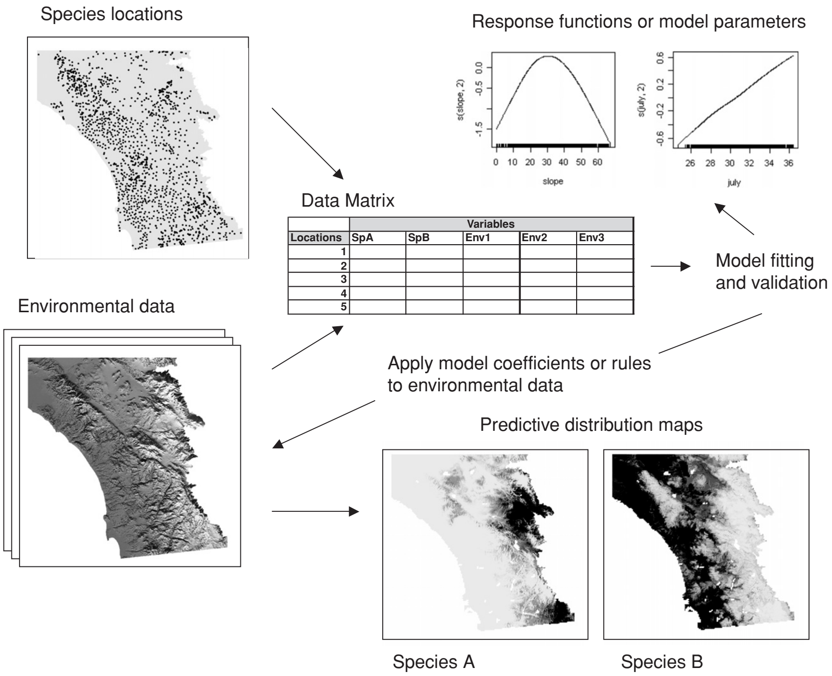

Fig. 1.1. The steps in species distribution modeling and predictive mapping. Spatially explicit (georeferenced) species occurrence data are linked with digital maps of environmental predictors and values of environmental predictors for species locations are extracted. A statistical, machine learning, or other type of model describing the relationship between species occurrence and environmental data is developed. The parameters or response functions are evaluated and these coefficients or decision rules are applied to environmental maps yielding spatial predictions of species distribution or habitat suitability.

for more advanced practitioners. There are a wide variety of applications of SDM ranging from basic to applied science. It is because of this that there are many choices that have to be made about what data and model to use. Given the range of subsequent decisions that will be made on the basis of SDMs, it is important to understand how they are constructed, using what types of applicable data, the assumptions behind them and their limitations. In the remainder of this chapter I will define some key terms, highlight some recent publications that illustrate the growing interest in SDM, outline the process of species distribution modeling along with the organization of the remainder of this book, and review some of the main uses of SDM.

1.2 What is in a name?

It seems a straightforward question, but what do species distribution models *describe*, and, literally, what do they *predict*? What variable is depicted in the resulting map? (This is discussed further in Chapter 4.) The terminology that has developed in this field was somewhat confusing when I reviewed it in 1995, with individual studies using different terms, and has grown even more complex. *Species distribution models*, of the kind discussed in this book, have been said to describe both (a) the species niche, and (b) the suitability of habitat to support a species.

1.2.1 Niche models

"Species niche model," "ecological niche model" or even "niche-theory model" are terms that have been used to describe SDM. These models have variously been described as estimating the fundamental (potential) niche, realized (actual) niche, the multivariate species niche (Rotenberry *et al.*, 2006), or, when conditioned only on climate variables, the "climatic niche" (Chapter 3 discusses this further).

The original climate niche modeling system (BIOCLIM) actually referred to the "climate profile" of a species or other entity (Busby, 1991). The term climate (or bioclimatic) envelope modeling (Heikkinen *et al.*, 2006) has been used even when envelope methods are not used (see Chapter 8) and when environmental predictors include variables other than climatic, especially when the models are applied to climate change questions. The maps resulting from the application of climatic niche models are often referred to as predictions of (species) geographical *range* (Graham *et al.*, 2004). Some authors even distinguish between ecological niche models, which they define as models of potential distribution, and species distribution models of actual distributions (Peterson *et al.*, 2008).

I prefer the term "species distribution model" to any of the variations on "niche model" (in spite of their widespread use) because SDM more accurately describes the modeling process and resulting model. Although niche theory strongly underpins SDM (Chapter 3), SDMs describe empirical correlations between species distributions and environmental variables. While these models should always be evaluated for "ecological realism," that is, consistency with ecological knowledge of limiting factors and species response curves (Chapter 3), the data rarely allow the true species niche to be fully specified or confirmed. Further, these models are frequently used to predict the geographical distribution of a species, rather than to study the characteristics of its distribution in environmental (niche) space.

1.2.2 Habitat suitability models

SDMs are also referred to as habitat suitability models (e.g., Hirzel *et al.*, 2006; Ray & Burgman, 2006; Hirzel & Le Lay, 2008), describing the suitability of habitat to support a species. When used to predict in geographical space, they have been called, among other things, "predictive habitat distribution models" (Guisan & Zimmermann, 2000) and "spatially explicit habitat suitability models" (Rotenberry *et al.*, 2006).

The concept of habitat suitability is closely related to the idea of a resource selection function from wildlife biology (Manly *et al.*, 2002) as noted by Boyce *et al.* (2002), and can be applied to plants and animals. A resource selection function (RSF) is any function (for example, from a statistical model) that is proportional to the probability of habitat use by an organism (Manly *et al.*, 2002). If a resource selection function is proportional to the probability of use, then a SDM could be said to predict the likelihood that an event (species) occurs at a location – that is, *the probability of species presence*. Mackey and Lindenmeyer (2001) presented a very interesting analytical framework for modeling the spatial distribution of terrestrial vertebrates which tend to be mobile and may not occupy all suitable habitat (van Horne, 1983). They refer to the spatial occurrence and abundance of a species as its "distributional behavior" (discussed further in Chapter 3).

A species distribution model, when applied to maps of environmental variables, is said to predict the species potential geographical distribution (potential occurrence at a location); the resulting maps have been called ecological response surfaces (Lenihan, 1993); biogeographical models of species distributions (Guisan *et al.*, 2006), spatial predictions of species distribution (Austin, 2002), predictive maps (Franklin, 1995), predictions of occurrence (Rushton *et al.*, 2004), or predictive distribution maps (Rodriguéz *et al.*, 2007). According to Meyer and Thuiller (2006), one of the most important applications of RSFs is mapping species distributions.

Further, the same methods and spatial modeling principles apply whether the locations of plants, animals, microbes, community properties (e.g., diversity), ecosystem properties (e.g., productivity), or other response variables are being correlated with the mapped distributions of those factors hypothesized to control their geographical patterns. Some theory other than the species niche concept (Chapter 3) would have to support such analyses, for example an ecosystem productivity model or community diversity theory. These analyses might also have to rely on different predictors than those typically used in SDM (Chapter 5). Spatial overlay of geographical patterns to examine the suitability of a location for a particular use forms the historical foundation of geographical information science (McHarg, 1969).

The description of SDM that I have used before (Franklin, 1995), "geographical modeling of biospatial patterns in the relation to environmental gradients," while comprehensive, is cumbersome. In this book I will use the terms *species distribution model*, and *predictive [distribution] map*, for simplicity and generality.

1.3 Heightened interest in species distribution modeling

"The quantification of . . . species–environment relationships represents the core of predictive geographical modeling in ecology" (p. 148) (Guisan & Zimmermann, 2000). Species distribution modeling has its roots in ecological gradient analysis (Whittaker, 1960; Whittaker *et al.*, 1973), biogeography (Box, 1981), remote sensing and geographic information science (for review see Franklin, 1995). It has recently experienced explosive growth in the scientific literature, and in practice, especially by governmental and non-governmental organizations charged with biological resource assessment and conservation at larger spatial scales. The level of activity has been greatly enhanced by the development of digital databases for natural history collections, making species location information from specimen records and other sources more widely available (Graham *et al.*, 2004), for example via the Global Biodiversity Information Facility (http://www.gbif.org/). In 1995, I wrote a "comprehensive" review of the state of the art that described only 30 studies. SDM has since been the subject of a number international symposia and workshops resulting in an edited book (Scott *et al.*, 2002) and special features (collections of papers) in journals:

- *Ecological Modelling*, 2002, **157** (2–3) and 2006, **199** (2);
- *Biodiversity and Conservation*, 2002, **11** (12);
- *Journal of Applied Ecology*, 2006, **43** (3);
- *Diversity and Distributions*, 2007, **13** (3).

It is a topic that is, or has been, featured often in *Ecography*, *Ecological Applications*, *Journal of Vegetation Science*, and in a number of other scientific journals. Pick up any recent issue and you will find several

SDM studies. For example, the October 2008 issue of the journal *Ecological Applications* featured a study distinguishing the factors affecting the establishment piñon–juniper woodland before, versus after, Euro-American settlement in western North America using topo-climatic predictors (Jacobs *et al.*, 2008). Another study in the same issue showed that improved predictions of marine predator (bottlenose dolphin) distributions could be made using models of prey (fish) distributions (Torres *et al.*, 2008). Species distribution modeling methods have even been used to map the human "ecological" niche during the last glacial maximum in Europe (Banks *et al.*, 2008), and the "biophysical environmental space" of wildfire (Syphard *et al.*, 2008; Parisien & Moritz, 2009). These are just a few of the many fascinating SDM studies to pass my desk.

Several review papers, essays and editorials have summarized advances in different aspects of SDM. These advancements include conceptual issues such as a general framework for species distribution modeling (Franklin, 1995; Guisan & Zimmermann, 2000; Mackey & Lindenmayer, 2001; Guisan & Thuiller, 2005; Elith & Leathwick, 2009), links to ecological theory (Austin, 2002, 2007; Hirzel & Le Lay, 2008), modeling methods (Guisan *et al.*, 2002, 2006; Pearce & Boyce, 2006), data and scale issues and statistical model selection (Rushton *et al.*, 2004), use of natural history collections data (Graham *et al.*, 2004), modeling ecological communities (Ferrier & Guisan, 2006), and the influence of spatial autocorrelation on models of species distributions (Miller *et al.*, 2007). Other reviews have focused on applications, including the use of SDMs in land management under uncertainty (Burgman, 2005), and their applicability to conservation planning (Ferrier *et al.*, 2002a, b; Rodriguéz *et al.*, 2007). Each of these topics will be explored in this book.

Further, growing numbers of government agencies and non-governmental organizations have implemented ambitious, large-scale species distribution modeling programs (Iverson & Prasad, 1998; Ferrier, 2002). These projects usually involve modeling the distributions of hundreds (even thousands) of individual species or ecological communities over large regions. Often, these efforts are extensive and influential, affecting regional and global conservation decision making, but they are not necessarily reported in the scientific literature. For example, NatureServe, a non-profit conservation organization that represents an international network of biological inventories (http://www.natureserve.org/), calls their program predictive (or element) distribution modeling and they are using it to model species distributions in South America. They have run training workshops and provided tutorial materials (for developing the statistical models) on their website (http://www.natureserve.org/

prodServices/predictiveDistModeling.jsp). The American Museum of Natural History, as part of their Remote Sensing and Geographic Information System Facility, has offered short courses in species distribution modeling for conservation biology (http://geospatial.amnh.org/). Clark Labs, in collaboration with Conservation International, an international non-governmental conservation organization, has developed specialized commercial GIS software to implement both species distribution models and project the impacts of land cover dynamics on biodiversity (http://www.clarklabs.org/). It is being used to model the distributions of 16 000 species in the Andes.

It is challenging to try and summarize the wealth of new information being produced in this exciting and dynamic field into a coherent narrative that can guide those who wish to use these methods in their work, but that is the goal of this book. Much recently published research in this area has focused on comparisons of different modeling methods and those comparisons have emphasized different measures of model performance or accuracy. While the findings of that important and useful work will be summarized, this book has a somewhat different emphasis. SDMs are empirical models and they start with data. I will first consider the qualities or characteristics of the data describing environment and species distributions in relation to the question being asked. This is in line with more recent studies that have emphasized the effect of species and environmental data properties on model performance. Trained as a geographer, I will emphasize the spatial nature of the data.

I will also emphasize how research informs the practice of SDM with reference to the published literature, as in the preceding paragraphs. Attribution is very important in science and scholarship, and I have tried to not only give credit where credit is due, but also provide signposts to where to look for those who want to find more detailed information about a particular topic.

1.4 What is species distribution modeling and how is this book organized?

This book is organized according to a framework for modeling species distributions presented by Austin (2002) that has three parts: the ecological, data, and statistical models (model here meaning framework or conceptual model). The ecological model includes the ecological theory applied or hypotheses 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" (Austin, 2002, p. 102). The

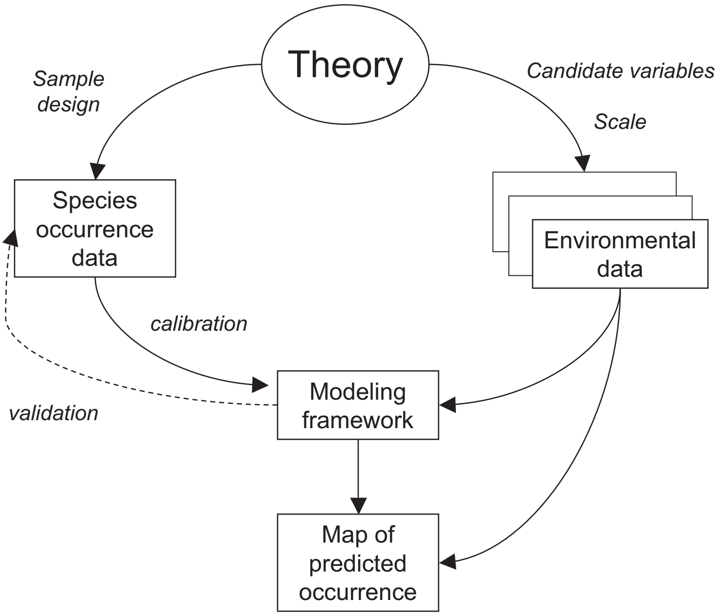

Fig. 1.2. Diagram showing the components of species distribution modeling. Biogeographical and ecological theory and concepts frame the problem, and identify the characteristics of the species and environmental data required to calibrate an appropriate empirical SDM and apply it to produce a map of predicted species occurrence or suitable habitat.

statistical model includes the choice of methods and decisions regarding implementation (calibration and validation).

The following elements are required for modeling and spatial prediction of species distributions and are described in the chapters of this book (Fig. 1.2):

- A theoretical or conceptual model of the abiotic and biotic factors controlling species distributions in space and time, and at difference scales, and the expected form of the response functions (Chapter 3);
- Data on species occurrence (location) in geographical space (a measure of presence, habitat use, abundance, or some other property) (Chapter 4), or expert knowledge about habitat requirements or preferences (Chapter 7);
- Digital maps of environmental variables representing those factors (or their surrogates) determining habitat quality, or correlated with it (Chapter 5). These are generally derived from remote sensing, from