Chapter 7
Species Distributions

7.1 Introduction

Understanding and predicting species distributions lies at the heart of ecology. Predictive models of species distributions are increasingly used in both basic and applied ecology to predict the effects of future climate change (Thomas et al. 2004), land-use change (Feeley and Silman 2010; Martin et al. 2013), species invasion (Peterson 2003; Elith et al. 2010; Jimenez-Valverde et al. 2011), agricultural suitability (Evans et al. 2010; Plath et al. 2016), best places for species reintroduction (Hirzel et al. 2004; Martinez-Meyer et al. 2006), identify new protected areas (Wilson et al. 2005), and to refine biodiversity inventories (Raxworthy et al. 2003).

Over the past two decades, there has been an explosion in the advancement and application of predictive distribution models (Guisan and Zimmermann 2000; Elith and Leathwick 2009; Renner et al. 2015). Species distribution models (SDMs), ecological niche models (ENMs), climate envelope models, and habitat suitability models (HSMs) all describe models that relate species distribution (occurrence or abundance) to the environment through the quantification of response surfaces (i.e., relationships of species distribution with environmental variables; Guisan and Zimmermann 2000; Guisan et al. 2017) (Fig. 7.1). Other related models include resource selection functions, occupancy models, and GAP models (Scott et al. 1993; Manly et al. 2002; Rodrigues et al. 2004; MacKenzie et al. 2006). These models are used for both inference on environmental relationships as well as prediction and projection, where the estimated functions are used to map distributions over space and time. These types of models have been developed in different sub-disciplines and each has a unique focus on the types of questions addressed, the scales at which questions are typically asked, and the specific types of data that are used. However, they all emphasize the relationship of species distribution with the environment.

Here, we describe the key concepts relevant to predicting species distributions, the types of data typically used, some common modeling algorithms, and illustrate how models are frequently evaluated. Our general goal is to illustrate how concepts,

R. Fletcher, M.-J. Fortin, *Spatial Ecology and Conservation Modeling*,
https://doi.org/10.1007/978-3-030-01989-1_7

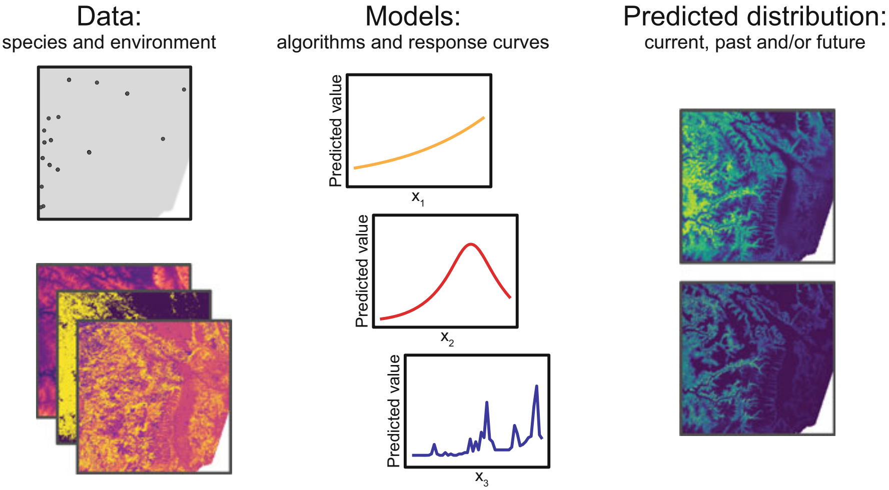

Fig. 7.1 A general framework of modeling species distributions. Data on species location are linked to spatial data on the environment with quantitative models. These models vary considerably in their assumptions about species responses to environmental gradients. With estimated response curves, species distributions are mapped in space and/or time. Modified from Guisan et al. (2017)

data and models are used to create maps of species distributions for addressing ecological questions and conservation problems. For more information regarding general species distribution concepts, see the excellent books by Franklin (2009), Peterson et al. (2011), and Guisan et al. (2017).

7.2 Key Concepts and Approaches

7.2.1 The Niche Concept

> No concept in ecology has been more variously defined or more universally confused than 'niche'. Nonetheless, the concept has become symbolic of the whole field of ecology. Real and Brown (1991)

Species distribution modeling generally relies on niche concepts for developing models (Austin 2002, 2007; Hirzel and Le Lay 2008). Most of the applications of niche theory is heuristic—that is to say that scientists tend to use general ideas that emerge from niche theory. Understanding the relevance of the niche concept is essential for building, interpreting, and applying distribution models to ecological, evolutionary, and conservation problems. Other theoretical developments have also been used in the context of predicting species distributions, such as habitat selection (Fretwell and Lucas 1970) and metapopulation theory (Pulliam 1988; Hanski and Ovaskainen 2003), but here we focus on developments related to niche theory.

7.2.1.1 A Brief History of the Niche Concept and a Plethora of Niches

The term *niche* was originally coined by Joseph Grinnell in the early 1900s. Grinnell was interested in the biogeography of birds and what limited their geographic range. He specifically considered the problem of spatial overlap in congener thrasher species in California (Grinnell 1917). His interpretation of the niche was largely based on the idea of species–environment relationships, emphasizing the role of habitat and behavioral adaptations as key components of a species' niche. For example, the California thrasher (*Toxostoma redivivum*) that he studied is well adapted to its environment, foraging in shrubs and having adaptive behaviors to reduce predation risk (e.g., camouflage).

In the 1920s–1930s, Elton advanced the niche concept, taking a different perspective. He emphasized the functional role of the species in its environment in relation to food and enemies, in which species could impact the environment through trophic interactions (Elton 1927). Elton focused on what a species does rather than where a species occurs, such that he focused on both the species response to, as well as the effect on, the environment. This perspective was quite distinct from Grinnell's perspective.

In the 1950s, Hutchinson took a quantitative perspective on the niche, considering it a "*N*-dimensional hypervolume where a species could persist" (Hutchinson 1957). In this way, an *N*-dimensional hypervolume reflects the idea that there are *N* environmental variables that are required for species persistence, each of which can be viewed as a different dimension in environmental space, and it is the intersection of suitable values of all *N* variables, or the hypervolume, that identifies the niche (Blonder et al. 2014). This work catalyzed the application of niche concepts—including niche breadth, niche overlap, and niche partitioning—by emphasizing measurable properties or dimensions of the niche. Hutchinson also distinguished between fundamental versus realized niches, where the *fundamental niche* was the environmental hypervolume in which a species could potentially persist (sometimes referred to as the physiological or potential niche), while the *realized niche* was a subset of this space where species actually occurred (Fig. 7.2). He assumed that the realized niche was smaller than the fundamental niche due to species interactions, particularly competition. This distinction is often made in the development of species distribution models (e.g., Guisan and Thuiller 2005). Hutchison (1957) defined the niche as a property of the species, not a property of the environment. As a consequence, for Hutchison, there were no "empty niches" in the world.

Pulliam (2000) and a seminal book by Chase and Leibold (2003) advanced the niche concept. Pulliam (2000) emphasized that dispersal limitation could result in many places that have environmental conditions that fall within a species' niche yet remain unoccupied. He also emphasized that the realized niche could in fact be larger than the fundamental niche in situations where species occurred in sink habitats (see Chap. 10). Chase and Liebold (2003) sought to unify niche concepts, integrating Eltonian and Hutchinsonian views in a common framework. They defined the niche

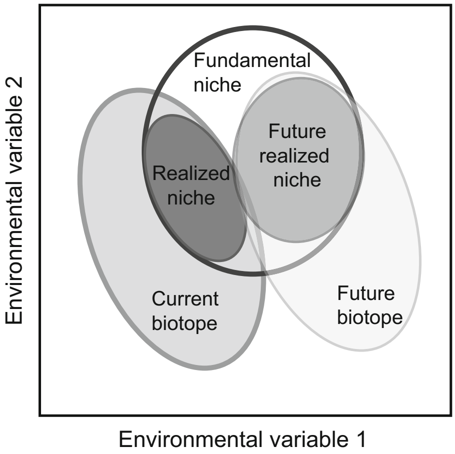

Fig. 7.2 Environmental gradients and the niche. Shown are two environmental variables relevant for the fundamental and realized niche of a species. The current biotope constrains the observed niche of a species, where the current biotope does not include all conditions of the fundamental niche. Changes in the biotope from environmental change causes a shift in the realized niche. Modified from Franklin (2009) and Williams and Jackson (2007)

as, "The joint description of the environmental conditions that allow a species to satisfy its minimum requirements so that the birth rate of a local population is equal to or greater than its death rate along with the set of per capita effects of that species on these environmental conditions" (Table. 7.1).

An important aspect of niches defined by Hutchison, and advanced by Pulliam (2000), Chase and Liebold (2003), and others, is that of fitness. From these perspectives, the niche embodies conditions where positive population growth occurs. Models of species distribution, in contrast, typically only use information on species occurrence or abundance (see below). Because species occurrence or abundance may not correlate with resource quality or positive population growth (Van Horne 1983; Schlaepfer et al. 2002; Robertson and Hutto 2006), using information on occurrence or abundance alone may not be sufficient for modeling the niche. That nuance may be fine under some situations, but it makes it clear that predicting species distributions may not be the same as predicting the niche. Indeed, there has been much debate regarding exactly what species distribution models really predict and how it relates to the niche concept (Franklin 2009; Araújo and Peterson 2012; Peterson and Soberón 2012).

7.2.1.2 Geographic Versus Environmental Space

When translating the niche concept to spatial models, a key distinction is geographic versus environmental space. This distinction was emphasized by Hutchinson's perspective on the niche, where he distinguished the niche from the geographic

Table 7.1 Common terms and definitions used in species distribution modeling

Term	Definition
Biotope	The community's environment (independent of a species).
Correlative distribution model	Predictive models that are based on response functions derived from species distribution and environmental factors.
Fundamental niche	The environmental hypervolume in which a species could potentially persist.
Mechanistic distribution model	Predictive models based on either experiments or known relationships of species with critical limiting factors, such as thermal tolerances of species.
Niche	The joint description of the environmental conditions that allow a species to satisfy its minimum requirements so that the birth rate of a local population is equal to or greater than its death rate along with the set of per capita effects of that species on these environmental conditions.
Realized niche	The environmental hypervolume in which a species occurs, or a subset of the fundamental niche where favorable biotic interactions occur.
Sample selection bias	Bias that can arise when samples of species distribution are nonrandom from the underlying distribution. Common in data collected opportunistically.
Species prevalence	How common a species is across the extent under consideration.

space that contains environmental variation, or the *biotope* (Whittaker et al. 1973; Colwell and Rangel 2009) (Fig. 7.2). A consequence of this distinction is that the niche is clearly an attribute of a species or population (one cannot have "empty niches"), which also helps to understand the differences between niches and habitats (see Chap. 8). Often we aim to make inferences on environmental space (e.g., functions that describe species responses to environmental conditions) and then we wish to map these responses in geographic space to make predictions or projections of species distributions (Fig. 7.1).

The current biotope and future biotope may vary considerably, and our understanding of niches is constrained by the biotope where species currently occur (Fig. 7.2). For instance, when using niche concepts to understand and predict the effects of climate change, current data on species niches may be insufficient because portions of the fundamental niche may not be expressed under current conditions (Williams and Jackson 2007). Similar problems arise with predicting the spread of invasive species using information on the environment in the native range of the species (Peterson 2003; Broennimann et al. 2007). Experiments can help partially resolve this issue (Buckley and Kingsolver 2012).

7.2.1.3 Limiting Factors and the Niche

Several factors can limit the dimensions of the niche (Araújo and Guisan 2006). Soberón categorized these factors as being one of three categories: abiotic, biotic, and movement-related limitations (Soberón 2007, 2010; Soberón and Peterson 2005). He visualized and interpreted these limiting factors using Venn diagrams and set theory, what are referred to a "BAM" diagrams (Biotic-Abiotic-Mobility diagrams; Fig. 7.3). Where these three factors intersect defines the current geographic distribution of a species. **A** captures favorable abiotic conditions and non-interactive variables ("scenopoetic variables") where the intrinsic growth rate of a species is positive, what has been termed the Grinnellian fundamental niche (James et al. 1984). **B** emphasizes the area where biotic interactions (sometimes referred to as Eltonian factors) allow for positive population growth. **M** represents the area that is accessible to organisms, that is, the colonizable area (Barve et al. 2011). In this context, the geographic expression of the *realized niche* has been described as the intersection of B and A, where conditions are suitable but movement limitations may or may not preclude species occurrence (Peterson et al. 2011; Soberón and Peterson 2005). Soberón argued that biotic factors are typically only relevant at fine spatial grains and thus can potentially be ignored in predicting broad-scale distributions and ranges of species (Soberón 2010; Busby 1991), termed the Eltonian Noise Hypothesis (Soberón and Nakamura 2009), although this conclusion is often debated (e.g., Wisz et al. 2013).

In general, the relative importance of these limiting factors may vary across species and across spatial scales (Pearson and Dawson 2003; Soberón and Peterson 2005). For example, Lira-Noriega et al. (2013) found that the importance of dispersal-

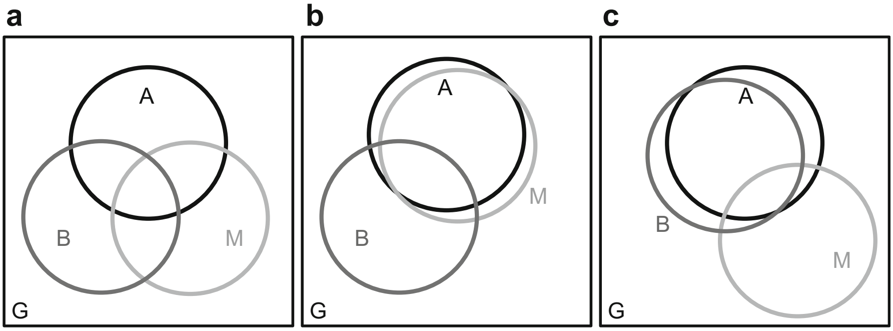

Fig. 7.3 BAM diagrams, illustrating the intersection of abiotic, biotic and movement-related limiting factors of relevance to the niche within a study region, G (or biotope). In this framework, A is considered the Grinnellian fundamental niche, which may or may not be occupied in a region depending on biotic interactions and movement limitations. In (**a**) similar overlap of limiting factors occurs. (**b**) At a fine spatial scale (small grain and extent), the fundamental niche is entirely accessible, but may not be fully occupied due to biotic interactions. (**c**) At a broad spatial scale (coarse grain and large extent), movement limitations may prevent colonization of some portions of the fundamental niche, while biotic interactions at a coarse grain have been hypothesized to have small effects on distribution. Modified from Soberón and Peterson (2005)

related constraints for distributions of mistletoe (*Phoradendron californicum*) varied as a function of spatial resolution of models, where dispersal-related constraints were more important at fine resolutions. It has also been argued that this framework helps illuminate differences in approaches and philosophies for modeling species distributions (Soberón and Nakamura 2009).

7.2.2 Predicting Distributions or Niches?

The focus on predicting and mapping distributions has led to a wide array of terms to describe such models and projections. Often these models are referred to as "ecological niche models," "environmental niche models," "habitat suitability models," or "species distribution models" (Franklin 2009; Peterson et al. 2011; Guisan et al. 2017). In this context, a common question arises: are these efforts actually modeling niches (Peterson and Soberón 2012)? One argument is that if the focus is on environmental space, rather than geographic space, these modeling efforts are more squarely in the vein of modeling the niche (Peterson and Soberón 2012). Yet as all modern concepts regarding the niche emphasize that it is the environmental space is when species can persist (Holt 2009), we suggest refraining from interpreting these models as that of the niche when only distribution information (and no demographic information) is used for model building. While such models can provide hypotheses regarding niches, they are best viewed as modeling distributions.

7.2.3 Mechanistic Versus Correlative Distribution Models

The vast diversity of species distribution models can be organized in several ways. Two useful properties include whether models are correlative (e.g., phenomenological) or mechanistic (i.e., process-based), and the types of response data that are used.

Correlative distribution models take information on species distribution, such as presence records, and relate them to environmental covariates, based on some sort of formal relationship. These models are typically phenomenological—models that describe or explain patterns without regard to underlying mechanisms. In contrast, *mechanistic distribution models* are typically based on either experiments or known relationships of species with critical limiting factors, such as thermal tolerances of species (Buckley 2008; Kearney and Porter 2009). It is often argued that mechanistic models may be more valuable when extrapolating model predictions to new places or times; however, formal comparisons between correlative and mechanistic models have revealed similar model performance in some situations (Buckley et al. 2010). Furthermore, there are strengths and limitations to both approaches and it has been argued that perhaps this is a false dichotomy in the characterization of distribution models (Dormann et al. 2012).

7.2.4 Data for Correlative Distribution Models

Correlative models can also be categorized based on whether presence-only, presence–absence (or detection-non-detection; MacKenzie et al. 2002), or count (abundance) data are used (Brotons et al. 2004; Lutolf et al. 2006; Potts and Elith 2006; Aarts et al. 2012). Presence-only data, or data where only a sample of presence locations are available (and no information is available on absence or abundance), are commonly used in correlative distribution modeling (Elith et al. 2006). There are many sources of such data, including museum and herbaria specimens, information from citizen science programs, and atlas programs (Graham et al. 2004). Presence-only data can be used in isolation or they can be compared to background points, sometimes called "pseudo-absences" to build distribution models. The latter approach has been shown to frequently produce more accurate species distribution models than using presence data alone (Elith et al. 2006). The value of using background points is that it provides information on the biotope and if presence locations reflect a non-random distribution of the underlying environment available to organisms (Fig. 7.4). Two challenges with using background points are determining the number of background points and their spatial distribution (VanDerWal et al. 2009; Barbet-Massin et al. 2012). Some studies have attempted to select background

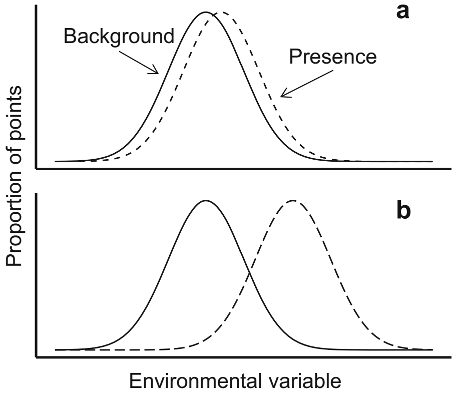

Fig. 7.4 The use of background or pseudo-absence points for presence-only modeling can provide relevant information on the biotope in the region for comparison to presence locations based on the difference between the environment at presence locations relative to the background locations. In (**a**) presence and background points have a similar distribution of environmental values, suggesting random distribution relative to the environmental gradients, whereas in (**b**) presence locations suggest a non-random distribution where the species is more likely to occur at high values of the environmental gradient

points that may be more likely to be considered absences based on using certain rules, such as only creating background points at minimum distances away from presence points; however, it is more common to simply generate randomly distributed background points. Renner et al. (2015) recently argued that many background points should be generated—more than commonly implemented in the literature—based on describing the use of background points in the context of inhomogeneous point process models (see below).

Presence-only data have the benefit of being plentiful across broad geographic areas. Furthermore, it is sometimes argued that such data circumvent the problem of false negatives in presence–absence data (i.e., recording an absence when in fact the species is present) (Guisan et al. 2007). Nonetheless, important limitations of such data include that there is often sample selection bias in opportunistic presence-only data and that the prevalence of the species is unknown. Both of these issues are valid concerns. *Sample selection bias* occurs when samples are a nonrandom sample from the region of interest, which often occurs in presence-only data when observations are more likely to be documented near easily accessible areas, such as near roads or urban areas (Kadmon et al. 2004; Loiselle et al. 2008; Phillips et al. 2009; McCarthy et al. 2012). Such bias can result in the identification of spurious environmental relationships and inaccurate predictions of distributions, in which models may provide predictions of sampling bias rather than underlying distributions. Unknown *species prevalence* arises because presence-only data do not provide information on how common the species is in the extent under consideration, because it is unclear if the presence-only samples reflect a small or large proportion of the underlying distribution. For instance, 30 presence records may be available for a species in a large study region, which could be because the species is rare and this number reflects the prevalence of the rare species, or it could be that the species is common and it was just inadequately sampled. This uncertainty leads to the conclusion that the probability of occurrence cannot be directly estimated with presence-only data (Yackulic et al. 2013) without making strong assumptions (Royle et al. 2012; Hastie and Fithian 2013). Instead, these models predict a relative measure of occurrence that is assumed to be proportional to the true probability of occurrence, similar to interpretations of resource selection functions relative to resource selection probability functions (see Chap. 8).

Presence–absence data, on the other hand, typically come from planned, standardized surveys. These types of data allow for formal modeling of the probability of occurrence of species (and can potentially account for observation errors and imperfect detection in the estimation of occurrence). These types of data are also thought to suffer less from sample selection bias. The rationale is that even when sampling may be biased across space or over time, because models are comparing occurrence observations to absence (or non-detection observations), effects of sample selection bias on estimated environmental relationships should be limited. Some have argued that because absence data may result from observation errors (false-negative errors), that it may be beneficial to only use presence data to help circumvent that problem (Guisan et al. 2007). However, in most situations imperfect absence data is still useful and can improve model predictions and interpretation (e.g., Brotons et al. 2004; Rota et al. 2011).

Count data are also sometimes used in distribution modeling (Guisan and Harrell 2000; Potts and Elith 2006), and are generally derived from planned survey data. Such data have the potential to provide abundance or density estimates. Count data provide greater information content and resolution in potential species–environment relationships (Cushman and McGarigal 2004); however, count data often require greater sampling intensity. Because distribution models are less frequently built with count data, for the remainder of this chapter we focus primarily on presence-only and presence–absence data.

7.2.5 Common Types of Distribution Modeling Techniques

We provide an overview on common modeling approaches for species distributions. Our summary is not comprehensive; rather we aim to emphasize very different techniques that capture the spectrum of variation in modeling algorithms. We illustrate envelope models (Pearson and Dawson 2003), the use of generalized linear and additive models (Guisan et al. 2002), regression trees and forests (Prasad et al. 2006; Elith et al. 2008), and Maxent (Phillips et al. 2006). We conclude by noting that many of these models can be derived more generally as inhomogeneous point process models (Renner et al. 2015), which may be helpful for better interpreting the relationships among model techniques.

These types of modeling algorithms are sometimes organized into three philosophically different approaches: profile methods, statistical models, and machine-learning algorithms. *Profile methods* are simple approaches that use environmental distances or similarity-based measures to relate environmental variability at presence locations to other locations across the region of interest. Some examples include envelope models (e.g., BIOCLIM), Mahalonobis distance, and DOMAIN (Carpenter et al. 1993; Rotenberry et al. 2006). *Statistical methods* are typically variants of linear models, such as generalized linear and additive models (Guisan and Zimmermann 2000) (see Chap. 6). In these approaches, a model is specified and then fit to the data via maximum likelihood or related techniques (e.g., ordinary least-squares). Statistical methods frequently focus on estimation of parameters and providing measures of uncertainty. *Machine-learning techniques* focus on identifying (and classifying) structure in complex data, often for situations where non-linearities and interactions are expected to occur, with the frequent goal of accurate prediction or classification (Olden et al. 2008). These philosophical distinctions can, however, be unclear, as some algorithms can be described from both a statistical and machine-learning perspective (e.g., Phillips et al. 2006; Elith et al. 2011).

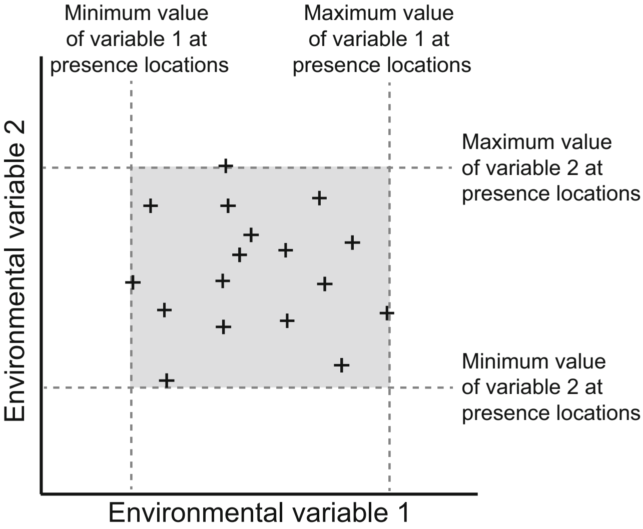

Fig. 7.5 Envelope models use information on either the observed minimum and maximum values of environmental factors at presence locations or quantiles of values (e.g., the 5%, 95% quantile). Pluses denote species occurrences and the grey box denotes the envelope

7.2.5.1 Envelope Models

Envelope models are presence-only models, wherein the distribution of environmental variation at presence locations is used to create an "envelope" of suitability. For example, the upper and lower quantiles of environmental covariates (e.g., 5–95% of elevation values) provide a means to create an envelope, where environmental conditions above or below those quantiles are deemed to be locations outside of the envelope (Fig. 7.5). There are many variations on this theme, but in general these approaches assume that all environmental variables considered are relevant, such that locations must be within the envelope of all variables.

The earliest applications of this approach focused on climatic variables and large-scale geographic range modeling. Busby (1991) developed software for this problem, BIOCLIM, which used climatic variables in a GIS to determine envelopes. More recent developments have attempted to gain more information out of envelope approaches by considering multivariate relationships among variables and through the use of similarity or kernel density measures to obtain relative measures of suitability.

7.2.5.2 GLMs and GAMs

In prior chapters, we introduced the use of generalized linear models (GLMs) and generalized additive models (GAMs). Both of these approaches are frequently applied to the problem of modeling species distributions. For distribution modeling, logistic models are typically used based on binary response data; however, these models are flexible and can also accommodate abundance response variables (Potts and Elith 2006). For presence-only data, presence points are typically contrasted to background points (Elith et al. 2006). While initial applications of this approach were somewhat ad hoc, this form of logistic regression can approximate more theoretically motivated inhomogeneous point process models (see below). See Chap. 6 for more detailed discussion of these methods.

Although GLMs have been widely used, a primary concern for their implementation in distribution modeling is the fact that may not adequately capture non-linear response functions, which are often emphasized in niche theory (Austin 2007). Because GAMs can accommodate non-linearity through the use of splines, they are frequently used as a logical extension of GLMs. Nonetheless, the types of non-linearity captured by GAMs (see Chap. 6) are less general than some other methods, such as Maxent and regression trees.

7.2.5.3 Regression Trees and Forests

An alternative to the generalized linear (and additive) modeling framework is the use of classification and regression trees (CART), also known as classification tree analysis (CTA), or recursive partitioning (RP). Classification trees work with data whose response variables are discrete, while regression trees work with continuous response variables. Like GAMs, they do not rely on a priori hypotheses about the relationship between independent and dependent variables. This method consists of recursive partitions of the values of predictors into groups that are as homogeneous as possible in terms of the response (Fig. 7.6). The tree is built by repeatedly splitting

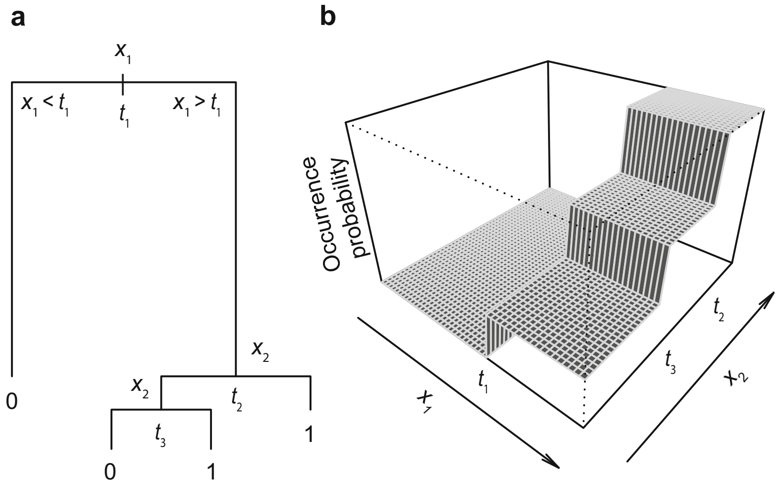

Fig. 7.6 A classification tree. (**a**) The splits of the tree and (**b**) how these splits result in species responses to environmental gradients. Modified from Elith et al. (2008)

the data, defined by a simple rule based on a single explanatory variable. At each split, the data are partitioned into two exclusive groups, each of which is as homogeneous as possible. A common approach is to grow a large tree and then prune it (i.e., reduce its size/complexity) by collapsing the weakest links identified through cross-validation and various indices (e.g., the "Gini" index; Breiman et al. 1984). The result can be thought of as a dichotomous tree that helps to classify locations of species occurrence. When the trees are short, they can be intuitive and visually appealing in terms of describing factors explaining distribution. As tree size grows, their interpretation can become more difficult.

Some advantages of this approach include the ability to easily handle non-linear relationships and interactions, outcomes are unaffected by monotonic transformations, trees are insensitive to outliers, and trees can accommodate missing data in predictor variables by using surrogates (Breiman et al. 1984; De'ath and Fabricius 2000). Nonetheless, CTA often performs poorly compared to GLMs, GAMs, and other models for species distribution modeling (Elith et al. 2008), in part because it has difficulty in modeling smooth functions and that CTA can be sensitive to small changes in the training (model building) data (Hastie et al. 2009; Guisan et al. 2017). However, two extensions of CTA—Boosted Regression Trees and Random Forests—are quickly being adopted because of their high predictive performance. We will focus on these methods, rather than CTA.

Random Forests and Boosted Regression Trees have gained popularity primarily because they typically provide high predictive accuracy relative to CTA and some other SDM algorithms (Elith et al. 2006; Prasad et al. 2006; Cutler et al. 2007). Rather than producing a single classification tree, these approaches are ensemble techniques that compile information from several models, using either "bagging" or "boosting." *Bagging* is a type of a bootstrap procedure, where several models are created through bootstrap sampling of the data (i.e., sampling with replacement) and predictions from models are combined in some way. *Boosting* uses sequential model development (a forward, stage-wise procedure), where with each iteration (sequence) there is an increasing emphasis on the training observations that are difficult to classify.

In Boosted Regression Trees, small, parsimonious trees are fit to the training data, with small trees sequentially added to the existing regression tree (Friedman 2002). The approach is stage-wise (rather than step-wise), meaning that with each iteration where new trees are added, the existing tree is left unchanged. The final model is a linear combination of many trees, analogous to a multiple regression model where each term is a parsimonious tree (Elith et al. 2008). There are two key parameters of interest when fitting a Boosted Regression Tree: the learning rate (or shrinkage parameter), which quantifies the contribution of each individual tree to the model, and the tree complexity, which controls the types of interactions considered. These parameters in combination will determine the number of trees used for predictions. Boosting has been shown to increase predictive abilities of models (Elith et al. 2006), reduce bias, and reduce variance in estimates, even when complex environmental relationships occur. For more on Boosted Regression Trees, see Elith et al. (2008).

Random Forests is a form of bagging, or bootstrap aggregation, where many trees are grown from bootstrap samples of the data, thereby producing a "forest" (Breiman 2001; Cutler et al. 2007). Predictions are made from each tree in the forest. Each tree gives a classification, such that each the tree "votes" for that class. The forest then chooses the classification having the most votes (across all the trees in the forest). Each tree is grown with the following steps. First, the training data are sampled with replacement (i.e., the data are bootstrapped). This sample is the training set for growing the tree. Second, for each node in the tree, n variables are selected at random out of N total variables (typically $n \ll N$) and are used to split the node in the tree. n is held constant during the forest growing, where each tree is grown to the largest extent possible (Breiman 2001). Accuracy and error rates are computed for each sample using the "out-of-bag" samples (those not used in the bootstrap sample) and are then averaged over all predictions. Some benefits of Random Forests include the following: (1) it can run efficiently on large datasets; (2) it can handle many explanatory variables and potential interactions; (3) it is argued to not over-fit; and (4) it can be used in several different types of problems (e.g., classification, survival analysis, clustering, missing value imputation) (Cutler et al. 2007).

7.2.5.4 Maximum Entropy

Maxent is a widely used approach for species distribution modeling, which uses the concept of maximum entropy (Phillips et al. 2006). Elith et al. (2006) provided a comprehensive analysis of the utility of different modeling algorithms for presence-only data and concluded that Maxent was one of the most useful algorithms. This result, coupled with available software that is relatively straightforward to implement, has led to widespread use of Maxent. In addition, it is frequently noted that Maxent is one of the only common distribution modeling algorithms designed specifically for presence-only data, because Maxent does not assume that background points are locations where the species does not occur (i.e., it is not assuming background points are absences), unlike the standard usage of GLMs, GAMs, and regression trees with presence–background data (but see Ward et al. 2009). As such, it might be particularly well suited for presence-only data.

The Maxent modeling framework can be described from several perspectives (Merow et al. 2013). In general, Maxent can be thought of as a log-linear model (Elith et al. 2011) and some parameterizations can be described more generally as an inhomogeneous point process model (Renner and Warton 2013; Phillips et al. 2017). The concept of maximum entropy states the best approximation of an unknown distribution is the one that is most spread out (or uniform), subject to some types of constraints (Franklin 2009). In this case, the constraints are derived from the expected value of the distribution estimated from the presence-only data. In its original formulation, Phillips et al. (2006) provided a geographic perspective regarding a Maxent probability distribution, where the Maxent distribution is equivalent to maximizing the likelihood of a Gibb's probability distribution, which can be written as:

$$p\Big(z(s_i)\Big) = \frac{\exp\Big(z(s_i)\lambda\Big)}{\sum_i \exp\Big(z(s_i)\lambda\Big)}, \tag{7.1}$$

where z is a vector of J environmental variables at locations s_i and λ is a vector of coefficients (Phillips et al. 2006). The numerator of Eq. (7.1) is a log-linear model, while the denominator is a normalization constant, such that $\Sigma\, p = 1$. Note that this latter aspect of the algorithm results in very small values for predictions at individual locations, but one can rescale p to make it more interpretable relative to other modeling algorithms (Elith et al. 2011; Phillips et al. 2017).

The Maxent package commonly used also includes other aspects of modeling that is not based on the idea of maximum entropy per se but rather general techniques employed in machine-learning modeling, such as model regularization and the use of "basis" functions or "features" to create non-linear response functions (Phillips et al. 2006; Phillips and Dudik 2008). In statistics, *model regularization* is an approach of shrinkage of parameter coefficients towards zero, which reduces potential overfitting of models (Tibshirani 1996). *Basis functions* or features are similar to the use of splines in GAMs, where features are an expanded set of transformations of the original covariates (Elith et al. 2011; Hefley et al. 2017). The practical difference between features used in Maxent and that of splines in GAMs is that Maxent can consider some functions that are not polynomial smoothers (e.g., cubic splines; see below). Maxent considers six types of features: linear, quadratic, product, threshold, hinge, and categorical (Fig. 7.7).

Because of these components to the Maxent program, there has been some confusion regarding *why* Maxent may be useful: is it because of the concept of

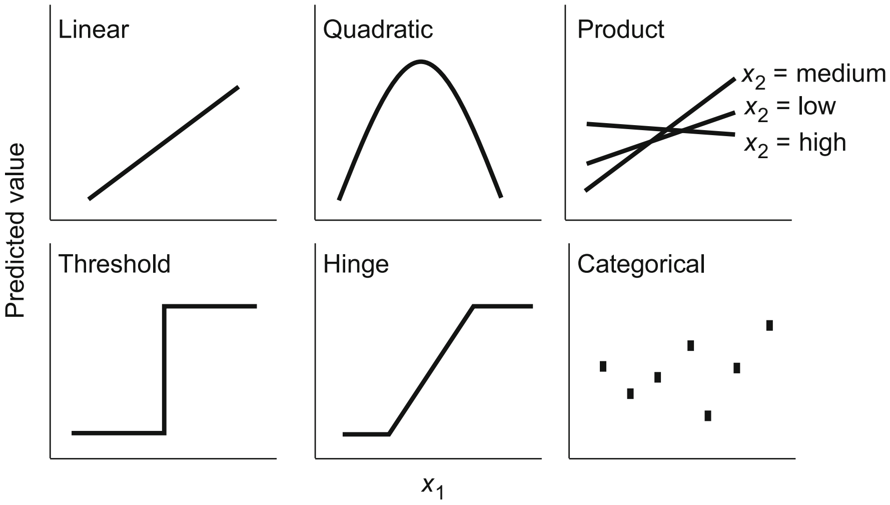

Fig. 7.7 Features considered by Maxent

maximum entropy or is it due to some of these other aspects? For instance, Gaston and Garcia-Vinas (2011) found that logistic regression that used a similar model regularization technique as Maxent (i.e., the "lasso"—least absolute shrinkage and selection operator) performed as well as Maxent, while logistic regression without regularization performed poorly. An important note: while Maxent can be run using presence–absence rather than presence–background (available) data, the underlying theory of the Maxent algorithm is based on presence-only data. Consequently, Maxent should not be used for presence–absence analysis of species distributions (Guillera-Arroita et al. 2014).

7.2.5.5 Point Process Models

It has been recently shown that several of the above modeling frameworks that focus on presence-only data can be derived more generally as spatial point process models (PPMs), including the use of Maxent, GLMs, GAMs, and Boosted Regression Trees (Aarts et al. 2012; Fitian and Hastie 2013; Renner and Warton 2013; Renner et al. 2015; Phillips et al. 2017). Previously, we have discussed point process models in the context of understanding spatial point patterns (Chap. 4). Here, the idea is that presence-only data can be viewed as point locations across a bounded region of interest, such that inhomogeneous point process models can describe the intensity (~density) of species, λ, in the region. The realization that many of the above SDM algorithms can be viewed as inhomogeneous point process models provides a unification of different modeling frameworks and it helps provide guidance for some recurring problems in distribution modeling (Warton and Shepherd 2010; Renner and Warton 2013; Phillips et al. 2017).

A point process is inhomogeneous when intensity varies across a region. Variation in intensity is captured by spatially explicit covariates by modeling intensity based on a log-linear relationship:

$$\log\lambda(s) = \alpha + \beta z(s), \tag{7.2}$$

where s is the species location. Consequently, PPMs are similar to Poisson regression (one type of generalized linear model; see Chap. 6), but the focus is on spatial locations of point occurrences rather than focus being on the point occurrences themselves (Fithian and Hastie 2013). In the likelihood of a point process model, there is a component that focuses specifically on estimating the background environmental conditions. This component can be approximated with background points (Berman and Turner 1992), referred to as "quadrature points" (because these points approximate the function that describes the background environment). Fithian and Hastie (2013) showed that by providing large weights to background points, logistic regression can approximate the inhomogeneous point process model and retrieve reliable parameter estimates of environmental relationships.

In a related way, Renner and colleagues (Renner and Warton 2013; Renner et al. 2015) showed how Maxent and other models can be derived as point process models in this framework. Why is this useful? By showing a common derivation, it illustrates the relationships among these techniques and better isolates exactly how they are different and implicit ways in which they are similar (e.g., some assumptions thought to not be relevant to Maxent but are relevant to GLMs may need to be reconsidered). In addition, this derivation provides important insight in some key aspects of species distribution modeling. For example, Warton and Shepherd (2010) and Renner et al. (2015) provided interesting discussions on how PPMs help clarify the role of background points and the number background points that should be included in analyses of presence-only data. Renner et al. (2015) emphasized that more background points should be used to estimate point processes than what is typically done in species distribution modeling. Also, the key for background points is that they should adequately capture the environment, such that they suggest that regular grids of points may be helpful, rather than random point generation.

The application of the PPM framework for species distribution modeling frequently only requires minor changes in model development. Renner et al. (2015) and Fithian and Hastie (2013) provided several examples of how these models can be implemented. In general, point process models can be fit in ways similar to other models, but typically more background points are used and presence and background points may be weighted differently (Fithian and Hastie 2013; Renner et al. 2015).

7.2.6 Combining Models: Ensembles

Because of the major differences in assumptions among modeling algorithms and their variable utility under different situations, ecologists have increasing used an "ensemble" approach to modeling (Araújo and New 2007). In a nutshell, ensemble models are typically (weighted) averages or related summaries of different model predictions (Fig. 7.8). For instance, in hurricane forecasting, ensemble predictions of hurricane paths are frequently used to get "consensus" predictions.

To make ensembles, we might take the median probability from a suite of models or take a weighted average, where the weights come from a measure of predictive accuracy (e.g., AUC or TSS). It is often argued that ensemble predictions can be more accurate than predictions from single models (Marmion et al. 2009). Nonetheless, care should be taken when using and interpreting ensembles, because some modeling algorithms are fundamentally predicting different currencies than others (e.g., envelope methods, GLM). For instance, profile methods typically predict environmental similarity while GLM-like models predicts (relative) probabilities of occurrence.

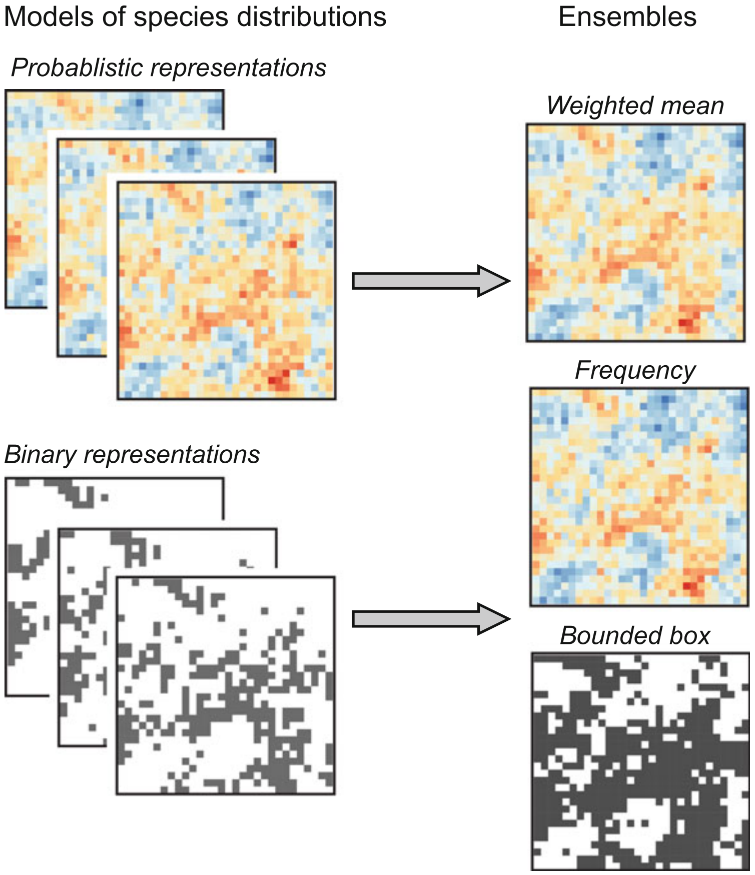

Fig. 7.8 Ensemble modeling integrates predictions from several models to make predictions of species distribution (e.g., occurrence). Ensemble predictions can be based on a variety of approaches. Shown are (weighted) averages of model predictions, frequencies of predicted occurrence based on binary summaries of model predictions, and the use of a bounded box, where at least one model predicts occurrence

7.2.7 Model Evaluation

Models can be evaluated in several ways. In the wildlife literature, there is a strong focus on model selection (e.g., AIC) (Burnham and Anderson 1998). Model selection can be very useful for contrasting the fit of models to data to compare hypotheses. However, model selection alone does not provide explicit information on the predictive performance of models, that is, the ability of models to predict to new locations in space or time, which is often of primary interest in distribution modeling (Hijmans 2012). In this way, predictions from models are frequently used for spatial interpolation (e.g., mapping species distributions), projections (e.g., evaluating alternative scenarios of expected climate change), and forecasting (e.g., making probabilistic predictions of species distribution in a new time or place).

To evaluate models, the primary approach is to build models with a portion of the data (sometimes called the "training" data) and then use the model to predict observations not used in the model building stage (sometimes called "test data" or "validation data"). This approach is commonly referred to as *external validation* (or *cross validation*) to distinguish it from *internal validation* (or *resubstitution*; see Fielding and Bell 1997), where a model is assessed based on predictions used to build the model. In general, external validation is thought to provide a much more honest assessment of the model performance than resubstitution. There are several approaches for partitioning data to be used for model building and testing. Prospective (independent) sampling, where new data are collected at different locations or time periods than for the data used in model training, is perhaps the most reliable approach but requires greater effort (Fielding and Bell 1997). In the absence of prospective sampling, K-fold partitioning is frequently used (Boyce et al. 2002). In K-fold partitioning, the data are split into K groups, or folds, and $K-1$ folds are used to predict the remaining fold that is left out of modeling training, such that K models are built and evaluated using each data point once as test data. This approach is an efficient way to split data for model evaluation. Folds are often created based on taking a random sample (without replacement) from the data for each fold, although this can result in test and training data being spatially interspersed (with the potential for spatial dependence in responses; see Chap. 6). Other approaches to creating folds include making spatial blocks of data ("K-fold block validation") (Wenger and Olden 2012) or stratifying random samples to ensure the same number of presence locations occur in each fold, such that the spatial distribution between training and testing data share similar characteristics (Hijmans 2012).

Once models are built and predictions are made onto new data, summary metrics are typically used to assess the predictive performance of models. The types of summary metrics and their utility depend on the type of response variable and evaluation data used in assessing models. We briefly summarize some common approaches with presence–absence models, presence-only models, and abundance models.

7.2.7.1 Evaluation with Presence–Absence Data

To evaluate predictions from presence–absence (or detection–non-detection data), we can either consider model discrimination or model calibration (Pearce and Ferrier 2000). *Model discrimination* assesses how well a model can tell the presences from absences (or background points) in the testing data set. In contrast, *model calibration* attempts to measure the agreement between predicted probabilities of occurrence and observed proportions of locations occupied in the testing data set.

Table 7.2 The confusion matrix

	Observed	
Predicted	Present	Absent
Present	a	b
Absent	c	d

Table 7.3 Common metrics derived from the confusion matrix (see Table 7.2 for constants used)

Metric	Equation
False positive rate (errors of commission)	$b/(b + d)$
False negative rate (errors of omission)	$c/(a + c)$
Sensitivity (True positive rate)	$a/(a + c)$
Specificity (True negative rate)	$d/(b + d)$
Correct classification rate	$(a + d)/N$
Prevalence	$(a + c)/N$
Kappa	$[(a + d) - (((a + c)(a + b)\ 1\ (b + d)(c + d))/N)]/$ $[N - (((a + c)(a + b) + (b + d)(c + d))/N)]$
True Skill Statistic	$a/(a + c) + d/(b + d) - 1$

Model Discrimination. For interpreting model discrimination, often the focus is on metrics that can be derived from the *confusion matrix*, or a summary table of predictions of presence–absence relative to observed presence–absence (Table 7.2). Typically, this matrix is obtained by truncating probabilistic predictions to 0/1 data, by selecting a threshold for truncating predictions. However, we note that Lawson et al. (2014) recently showed that the use of the confusion matrix need not require truncating predictions. There are several metrics that can be derived from the confusion matrix (Fielding and Bell 1997), including metrics that focus on certain types of errors in predictions (e.g., false positive or false negative errors), or overall model predictive accuracy (e.g., the correct classification rate). Here we focus on two metrics commonly used in distribution modeling: Kappa and the True Skill Statistic.

Kappa is a commonly used metric that expresses the agreement not obtained randomly between two qualitative variables. Kappa is a popular metric because it takes into account both omission and commission errors (Table 7.3). It is also less problematic than some simpler metrics taken from the confusion matrix, such as the correct classification rate (CCR) (Table 7.3), which can give a misleading interpretation of model performance because high CCR can occur when models predict all presences or all absences for common or rare species, respectively.

The True Skill Statistic (TSS), sometimes called the Hanssen–Kuipers Skill Score, has been traditionally used for assessing the accuracy of weather forecasts. TSS is typically defined as: sensitivity + specificity − 1. Like Kappa, TSS takes into account both omission and commission errors, as well as successes as a result of random guessing. It ranges from −1 to +1, where +1 indicates perfect agreement and values of zero or less indicate a performance no better than random. However, in contrast to Kappa, TSS is less affected by species prevalence (see Allouche et al. 2006). TSS is also thought to not be affected by the size of the validation set. TSS is a special case of Kappa when the proportions of presences and absences in the validation set are equal.

A common question pertains to how thresholds should be set for defining the confusion matrix. There are several approaches that have been used. Thresholding can be based on a general cutoff (e.g., predicted probability = 0.5), the prevalence of species in the training data, or more complex approaches, such as searching for the threshold that maximizes kappa or some other evaluation metric (Liu et al. 2005, 2013). Simple measures, such as using the prevalence (i.e., the proportion of sites occupied in the training data) can be useful (Liu et al. 2005). Liu et al. (2013) recommended searching for the value that maximizes the sum of specificity and sensitivity. In some cases, the type of error might matter (e.g., false positive or false negative rates may be more problematic in applications) and can be considered in this decision-making process (Fielding and Bell 1997).

Another popular metric for model discrimination is the Area under the Receiver Operating Characteristic (ROC) Curve (AUC), a curve representing the relationship between the false positive fraction (1 − specificity) and the sensitivity (true positive rate) for a range of thresholds. Good model performance is characterized by a curve that maximizes sensitivity for low values of (1 − specificity), that is, when the curve passes close to the upper left corner of the plot. The area under this curve (AUC) measures model discrimination. An AUC value of 0.5 can be interpreted as the model performing no better than a random prediction, with scores approaching 1 indicating progressively better performance. A value of 0.8 for the AUC means that for 80% of the time, a random selection from presence locations will have a prediction greater than a random selection from the absence locations (Fielding and Bell 1997). Thus, it is a rank-based discrimination metric and has a formal relationship to a Wilcoxon sign test. This metric is popular in part because is not dependent on using a threshold. AUC is widely used, but it is not without criticism (Lobo et al. 2008; Peterson et al. 2008). Some known issues with AUC is that it can vary depending on the spatial extent considered, where a larger extent tends to increase AUC. Because of this sensitivity, AUC can be misleading when compared in absolute terms across studies (although within an investigation it may be comparable among model algorithms). This criticism is relevant to other performance metrics as well. It is also frequently argued that the entire range considered by AUC is not biologically meaningful (Lobo et al. 2008). Finally, AUC was developed for presence–absence types of data. Its application to presence-only data should be used with caution.

Model Calibration. Model calibration is an important way to evaluate presence–absence models, where predicted probabilities are contrasted to observed proportion presences (or observed probabilities) in testing data. For example, a model could have good discrimination and yet consistently under (or over) predict the probability of occurrence. Such bias could be problematic when applying models to conservation problems.

Model calibration can be accomplished through two general approaches. First, a common way to interpret how well as model is calibrated is through the use of calibration plots. In this approach, predicted probabilities of occurrence and observed proportions of sites occupied are contrasted. To do so, often validation data are pooled based on predicted probabilities. By pooling observations from validation data, the proportion of locations occupied can be calculated (rather than relying solely on binary data). This is akin to some types of goodness-of-fit tests in statistics. These plots can be compared qualitatively or more quantitatively, such as comparing regression lines (e.g., intercepts, slopes) fit through different calibration plots (Guisan et al. 2017). Second, in addition to calibration plots, some metrics can be used, such as metrics that focus on the variation explained, error, and likelihoods (Lawson et al. 2014). In particular, the cross-validated log-likelihood and/or deviance ($-2 \times$ log-likelihood) can be calculated as a measure of model calibration (Lawson et al. 2014; Fithian et al. 2015), which have a strong foundation in statistical theory. In this context, the cross-validated log-likelihood ($\mathrm{LL}_{\mathrm{cv}}$) is defined as:

$$\mathrm{LL}_{\mathrm{cv}} = \sum_i \log(p_i y_i + (1 - p_i)(1 - y_i)), \tag{7.3}$$

where p_i is the predicted probability for observation i and y_i is the observed presence or absence of the species in the test data.

7.2.7.2 Evaluation with Presence-Only Data

Evaluating presence-only models can be challenging. When test data are presence–absence, the approaches mentioned above are frequently employed (Elith et al. 2006; Hijmans 2012), although care should be taken because of the subtle differences in model training and testing data. However, when test data are also presence-only, the approaches for presence–absence data should not be used. In such situations, evaluations should be based only on presence locations (and not the background or pseudo-absence locations) (Hirzel et al. 2006). One popular index is the Boyce Index (Boyce et al. 2002). The rationale of this index is to compare the predicted frequency of suitability values at evaluation points for a b classes (where b are bins of suitability; e.g., 0.0–0.2, 0.21–0.4, etc.) to the expected frequency of points based on a random distribution of points across the study area. This approach has been extended to reduce the sensitivity of bin classes on observed outcomes (Hirzel et al. 2006). Phillips and Elith (2010) also provided an extension of calibration plots for presence-only data.

7.2.7.3 Evaluation of Abundance (Count) Responses

Evaluating non-binary responses (e.g., abundance data) is, in many ways, more straightforward that evaluating models based on binary data. In these cases, no transformation of predictions is needed, unlike models based on binary responses.

These approaches typically focus on how well models are calibrated rather than discrimination. Common statistics include the root mean squared error, the coefficient of determination (R^2), and correlation coefficient (Potts and Elith 2006). Root mean squared error, RSME, is defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - y_i)^2}, \tag{7.4}$$

where p_i is the prediction for observation i and y_i is the observed value. In addition, statistics such as the deviance or cross-validated log likelihood can be used.

7.3 Examples in R

We illustrate the process of fitting species distribution models to presence-only data. Our goals are to contrast common modeling techniques and illustrate how models can be interpreted and evaluated. We also illustrate how different types of model evaluation can alter the conclusions regarding the utility of species distribution models.

7.3.1 Packages in R

In R, there are several libraries that can be used for species distribution modeling. Four common "wrapper" packages include `dismo` (Hijmans et al. 2017), `sdm` (Naimi and Araújo 2016), `ecospat` (Di Cola et al. 2017), and `biomod2` (Thuiller et al. 2016). These packages call other packages to perform a variety of species distribution models, including all those mentioned above and several others. Each of the models considered in `dismo` and `biomod2` could be implemented with other packages in R. For the purposes of illustration, here we will largely use individual packages because this provides greater flexibility and transparency in model development. We will also use `dismo` for implementation of some models not available in other packages. We use the `PresenceAbsence` package for model evaluation, which has a comprehensive set of evaluation metrics (Freeman and Moisen 2008), but several other packages can also evaluate models.

7.3.2 The Data

We will return to the data used in Chap. 6 on spatial regression for illustrating species distribution modeling techniques: the Northern Region Landbird Monitoring

Program (Hutto and Young 2002). In this monitoring program, sampling locations consisted of point counts (100-m radius) along a transect (10 points/transect; transects were approximately 3 km long), with transects randomly selected within USFS Forest Regions across Montana and Idaho. These points were also resampled over time (temporal repeated measures), although we will not consider these temporal repeated measures here. We will subset data to consider presence-only observations to illustrate presence-only modeling, but we will use presence–absence data for model evaluation, similar to prior syntheses on presence-only modeling techniques (e.g., Elith et al. 2006).

We again focus on the varied thrush. McCarty et al. (2012) modeled the occurrence of several species in this region, including the varied thrush. The varied thrush is a species of conservation interest, in part because it has declined in the region over the past 30 years (see Chap. 6), and it is considered an "interior" and "old-growth" species (Brand and George 2001; Betts et al. 2018). McCarty et al. (2012) considered the following covariates: canopy cover, the presence of mesic forest, elevation, and mean annual precipitation (see also George 2000). We consider each of these factors. Original GIS layers for canopy cover and mesic forest were 15-m resolution digital land-cover maps developed by the United States Forest Service Northern Region Vegetation Mapping Program (USFS R1-VMP), using Landsat TM imagery and aerial photography (Brewer et al. 2004). McCarty et al. (2012) used a Principal Components Analysis (PCA) to reduce the number of canopy cover variables from three to two. One principal component reflected a linear gradient of canopy cover, which we use here, whereas the other component reflected a non-linear gradient (high factor loadings on intermediate categories of canopy cover). We consider the proportion of mesic forest within a 1-km buffer. The 1-km landscape scale was chosen on the basis of other investigations in this region that showed strong correlations of avian distribution at this scale (Tewksbury et al. 2006; Fletcher and Hutto 2008), although other scales could be considered to best determine the scale of effect (see Chap. 2). Elevation was derived from a 30-m resolution Digital Elevation Model. Prior to analysis, all GIS layers were aggregated to a 200-m resolution, reflecting the grain of the sampling unit (100-m-radius point counts). Mean annual precipitation data come from the PRISM Climate Group at Oregon State University (http://www.prismclimate.org).

7.3.3 Prepping the Data for Modeling

There are several steps to prepping data for distribution modeling, depending on the data sources. In particular, working with opportunistic data often requires vetting observation and collating information to create relevant data frames for modeling purposes. See Di Cola et al. (2017) for more on these issues.

We first load the response data and subset the data based on presence–absence, as well as the *x*-*y* coordinates for presence–absence locations, to allow for simple extraction for modeling. There are two sources of data we consider. The first comes from the entire region collected in 2004 (`vath.data`). The second we consider as independent (prospective sampling) validation data (`vath.val`) collected in the region in 2007–2008 at a subset of points considered in 2004.

```
> vath.data <- read.csv(file = "vath_ 2004.csv", header = T)
> vath.val <- read.csv(file = "vath_VALIDATION.csv", header = T)

#subset to presence-only / absence-only
> vath.pres <- vath.data[vath.data$VATH == 1,]
> vath.abs <- vath.data[vath.data$VATH == 0,]
> vath.pres.xy <- as.matrix(vath.pres[,cbind("x", "y")])
> vath.abs.xy <- as.matrix(vath.abs[,cbind("x","y")])

#validation
> vath.val.pres <-
  as.matrix(vath.val[vath.val$VATH == 1, cbind("x", "y")])
> vath.val.abs <-
  as.matrix(vath.val[vath.val$VATH == 0, cbind("x","y")])
> vath.val.xy <- as.matrix(vath.val[,cbind("x","y")])
```

Next, we will load raster grids that contain relevant spatial information on the covariates we will consider (Fig. 7.9).

```
> library(raster)
> elev <- raster("elev.gri") #elevation layer (km)
> canopy <- raster("cc2.gri") #linear gradient from PCA
> mesic <- raster("mesic.gri") #presence of mesic forest
> precip <- raster("precip.gri") #mean precip (cm)

#check maps
> compareRaster(elev, canopy)

##
[1] TRUE

> compareRaster(elev, mesic)

##
Error in compareRaster(elev, precip) : different extent
```

In this situation, these maps do not align because they are of slightly different resolutions and extent, where the mesic forest is a resolution of 210 × 210 m while the others are 200 × 200 m. The elevation and canopy layers have the same extent, but the others are slightly different. As a consequence, we cannot create a raster stack (or brick) of these data. To rectify this problem, we resample the precipitation and mesic forest layers to be consistent with the elevation and canopy layers. Note that we use the "`ngb`" method for `mesic`, a categorical (binary) variable, and "`bilinear`" method for `precip`, a continuous variable.

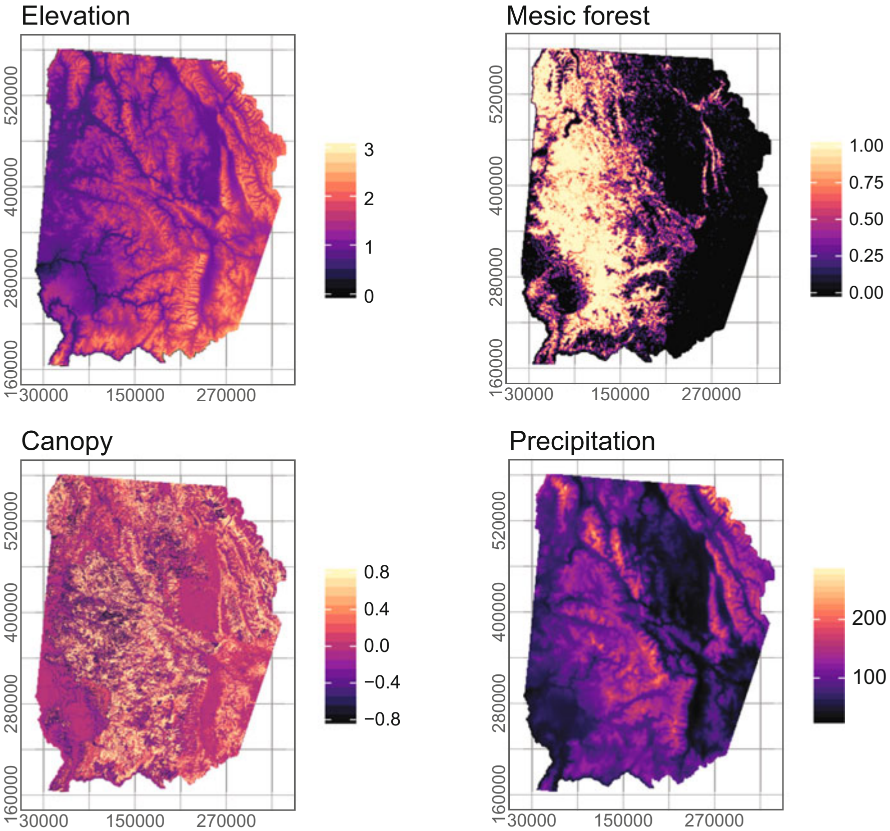

Fig. 7.9 Explanatory variables considered in model building, including elevation (in km), canopy cover (a derived metric taken from a Principal Components Analysis), the percent of mesic forest cover within 1 km of each location, and mean precipitation (cm)

```
#resample to align layers
> mesic <- resample(x = mesic, y = elev, "ngb")
> precip <- resample(x = precip, y = elev, "bilinear")

#crop to same extent
> mesic <- mask(mesic, elev)
> precip <- mask(precip, elev)

> compareRaster(elev, precip, canopy, mesic)

##
[1] TRUE
```

This resampling and masking aligns the raster data. Before we create a raster stack, we also add a larger scale covariate for mesic forest: the proportion of mesic forest in the surround 1 km.

```
#make 1 km wet forest
> fw.1km <- focalWeight(mesic, 1000, 'circle')
> mesic1km <- focal(mesic, w = fw.1km, fun = "sum", na.rm = T)
```

We can now create a raster stack of the environmental covariates (Fig. 7.9).

```
> layers <- stack(canopy, elev, mesic, mesic1km, precip)
> names(layers) <- c("canopy", "elev", "mesic", "mesic1km", "precip")

> plot(layers)
> pairs(layers, maxpixels = 1000)
```

Because mesic and mesic1km are highly correlated, we only consider mesic1km in further modeling. We can use the dropLayer function to remove that layer from the raster stack:

```
> layers <- dropLayer(layers, 3)
```

We can generate background points in several ways. The dismo package includes the randomPoints function for generating random points without replacement. For distribution modeling, we may want to generate points without replacement, because sampling with replacement would potentially create duplicate records (but see Renner et al. 2015). In addition, the raster package has the sampleRandom and sampleRegular functions, which can also generate availability points (without replacement). The randomPoints and sampleRandom functions are similar, but there is one key difference in the context of distribution modeling: the randomPoints function allows the user to also provide the presence points and, if so, it will not generate available points that those locations. We will illustrate the use of this package, generating 2000 background points. We choose this number for computational purposes only. In practice, we may want to increase this number substantially (Renner et al. 2015), but 2000 should be sufficient for illustration here.

```
> library(dismo)
> back.xy <- randomPoints(layers, p = vath.pres.xy, n = 2000)
> colnames(back.xy) <- c("x", "y")
```

With these locations and the points we identified above for presence and validation data, we extract covariate values at each point with the extract function, remove potential NAs (where random points were generated but not all environmental data occur), and link them into a single data frame:

```
> pres.cov <- extract(layers, vath.pres.xy)
> back.cov <- extract(layers, back.xy)
> val.cov <- extract(layers, vath.val.xy)

#link data
> pres.cov <- data.frame(vath.pres.xy, pres.cov, pres = 1)
> back.cov <- data.frame(back.xy, back.cov, pres = 0)
> val.cov <- data.frame(vath.val, val.cov)

#remove any potential NAs
> pres.cov <- pres.cov[complete.cases(pres.cov),]
> back.cov <- back.cov[complete.cases(back.cov),]
> val.cov <- val.cov[complete.cases(val.cov),]

> all.cov <- rbind(pres.cov, back.cov) #combine data
```

These data can now be used with a variety of modeling techniques.

7.3.4 Contrasting Models

7.3.4.1 Envelopes

Envelope models can be readily fit in the dismo package. In these models, we only use the presence locations. To create the envelope, the bioclim function in dismo calculates the percentiles of observed environmental covariates at presence locations and the values of covariates at each location on the map are compared to these percentiles. The closer the value of the location to the median value of a covariate at presence locations, the more suitable that location is deemed to be. Then, the minimum similarity value across covariates is used (analogous to Liebig's Law of the Minimum; Austin 2007). In our dataset, the envelope can be calculated as:

```
> bioclim.vath <- bioclim(layers, vath.pres.xy)
```

Here, the model will consider all covariates in the layer stack. We can plot the environmental variation at (~envelopes) the presence locations and produce a predictive map from this model:

```
#envelope plots
> plot(bioclim.vath, a = 1, b = 2, p = 0.85) #canopy-elev 85%
> plot(bioclim.vath, a = 1, b = 2, p = 0.95) #canopy-elev 95%
> plot(bioclim.vath, a = 1, b = 3, p = 0.85) #canopy-mesic 85%
```

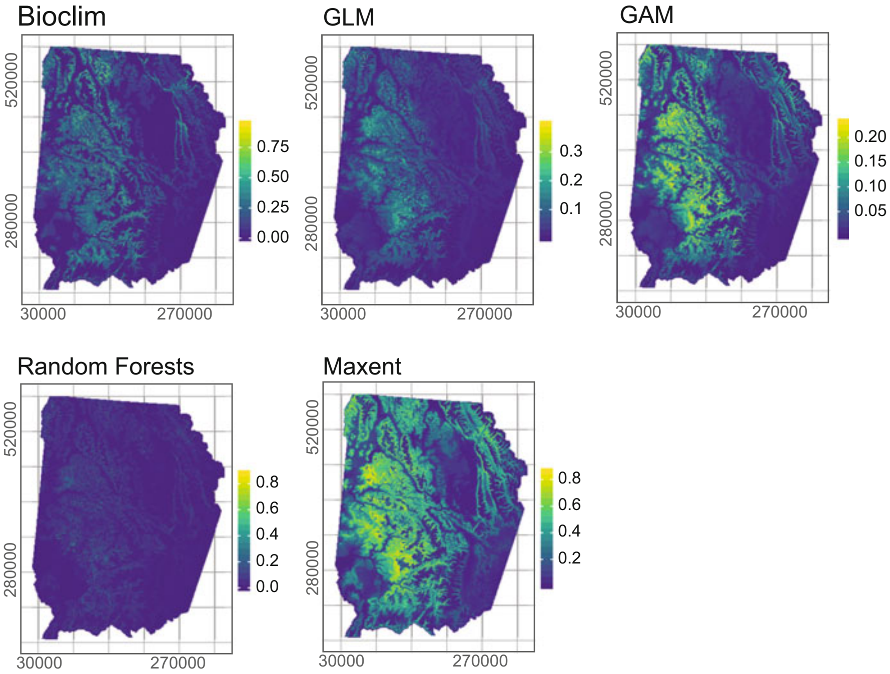

Fig. 7.10 Predictive maps taken from several distribution modeling techniques

```
#map it
> bioclim.map <- predict(layers, bioclim.vath)
> plot(bioclim.map, axes = F, box = F, main = "bioclim")
```

This map (Fig. 7.10) reflects the similarity of locations to environmental covariates at presence locations. It is scaled such that a value of 1 would be locations that have the median value of all covariates considered, while a value of zero would reflect locations where at least one covariate is outside the range of environmental covariates at presence locations.

While this model is simple in form, it illustrates the extent to which locations fall within the environmental variation of observed locations. Note that in doing so, it may often over-predict distributions.

7.3.4.2 GLMs and GAMs

Generalized linear models (GLMs) and generalized additive models (GAMs) are frequently used in distribution modeling. In these cases, presence–absence or presence–background data are used (Fithian and Hastie 2013). We fit a simple GLM (see Chap. 6 for more on GLMs and spatial regression models):

```
> glm.vath <- glm(pres ~ canopy + elev + I(elev^2) + mesic1km +
  precip, family = binomial(link = logit), data = all.cov)

> summary(glm.vath)

##
Call:
glm(formula = pres ~ canopy + elev + I(elev^2) + mesic1km + precip,
 family = binomial(link = logit), data = all.cov)

Deviance Residuals:
 Min 1Q Median 3Q Max
−0.8053 −0.3377 −0.2130 −0.1274 3.5746

Coefficients:
 Estimate Std. Error z value Pr(>|z|)
(Intercept) −12.186128 2.001925 −6.087 1.15e−09 ***
canopy 0.655128 0.282635 2.318 0.02045 *
elev 13.207998 3.251465 4.062 4.86e−05 ***
I(elev^2) −5.477279 1.293859 −4.233 2.30e−05 ***
mesic1km 1.127415 0.376421 2.995 0.00274 **
precip 0.011051 0.004529 2.440 0.01468 *
- - -
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

 Null deviance: 773.28 on 2093 degrees of freedom
Residual deviance: 667.91 on 2088 degrees of freedom
AIC: 679.91

Number of Fisher Scoring iterations: 8
```

In this model, we consider linear relationships for all covariates, except elevation, which we allow to be a non-linear (quadratic) relationship (`I(elev^2)`) because we expect that thrushes may be most likely to occur at moderate elevations (see Chap. 6). We do not consider model selection here, but model selection could be performed manually with a variety of packages, such as `MuMIn` (Barton 2018). This model suggests that there is a strong non-linear effect of elevation and a linear, positive effect of mesic forest within the surrounding 1 km. The other two covariates show weaker, positive linear relationships.

We can make a predicted map with:

```
> glm.map <- predict(layers, glm.vath, type = "response")
```

In this case, we specify `type ='response'` to make predictions on the probability scale. Otherwise, predictions would be on the link scale (here, the logit scale).

Generalized additive models can be fit with a few packages; here we illustrate the use of the `mgcv` package (Wood 2006). The default approach in `mgcv` is to optimally determine the number of knots via generalized cross-validation and to use thin-plate splines as a smoother. In this syntax, the `s()` function specifies that a spline will be applied to a covariate.

```
> library(mgcv)
> gam.vath <- gam(pres ~ s(canopy) + s(elev) + s(mesic1km) +
  s(precip), family = binomial(link = logit), data = all.cov,
  method = "ML")

> summary(gam.vath)

##
Family: binomial
Link function: logit

Formula:
pres ~ s(canopy) + s(elev) + s(mesic1km) + s(precip)

Parametric coefficients:
 Estimate Std. Error z value Pr(>|z|)
(Intercept) −4.068 0.252 −16.14 <2e−16 ***
- - -
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
 edf Ref.df Chi.sq p-value
s(canopy) 1.000 1.000 4.373 0.03651 *
s(elev) 3.157 3.997 23.796 9.28e−05 ***
s(mesic1km) 1.000 1.000 1.550 0.21316
s(precip) 4.403 5.226 19.671 0.00158 **
- - -
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) = 0.0709 Deviance explained = 17.3%
-ML = 335.55 Scale est. = 1 n = 2094
```

Results from this default GAM are generally similar to the GLM. We can tune the GAM by manually setting the number of knots (see Chap. 6, Fig. 6.5), requesting a different type of smoother function, or by allowing for potential interactions between predictor variables. We illustrate examples of each of these types of tuning. Also, note that we could include linear, rather than smoother terms, to the model by removing the `'s'` command around covariates. First, we specify the number of knots used manually, for example:

```
> gam.vath.knot3 <- gam(pres ~ s(canopy, k = 3) + s(elev, k = 3)
  + s(mesic1km, k = 3) + s(precip, k = 3), method = "ML", family =
  binomial(link = logit), data = all.cov)
```

```
> gam.vath.knot6 <- gam(pres ~ s(canopy, k = 6) + s(elev, k = 6)
  + s(mesic1km, k = 6) + s(precip, k = 6), method = "ML", family =
  binomial(link = logit), data = all.cov)
```

As the number of knots increase, the complexity of the smoother increases. Note that we also ask for model fitting with maximum likelihood (`method="ML"`), which allows us to make formal comparisons among models using model selection criteria. We can incorporate the potential for interactions between smoothers using a "tensor" product term. Tensor product smoothers address the potential for capturing interactions among variables that can be on different units of measurement (Wood 2006). They can be incorporated as:

```
> gam.vath.tensor <- gam(pres ~ te(canopy, elev, precip,
  mesic1km), family = binomial(link = logit), method = "ML", data
  = all.cov)
```

Finally, we can contrast thin-plate spline smoothers (the default in `mgcv`) to other smoother functions, such as a cubic spline ('`cr`'):

```
> gam.vath.cr <- gam(pres ~ s(canopy, bs = "cr") + s(elev, bs =
  "cr") + s(mesic1km, bs = "cr") + s(precip, bs = "cr"), family =
  binomial(link = logit), method = "ML", data = all.cov)
```

The relationships modeled when altering the number of knots and the smoother do not change much at all in this example (see Fig. 7.11 for some example plots). Overall, this model refines our understanding relative to the GLM, suggesting that varied thrush respond non-linearly to precipitation gradients in addition to elevation. This model tuning can be formally evaluated with model selection criteria, such as AIC:

```
> AIC(gam.vath, gam.vath.knot3, gam.vath.knot6, gam.vath.tensor, gam.
  vath.cr)
```

```
##
 df AIC
gam.vath 12.2 663.9
gam.vath.knot3 7.7 670.3
gam.vath.knot6 10.4 662.0
gam.vath.tensor 25.8 658.1
gam.vath.cr 12.9 663.9
```

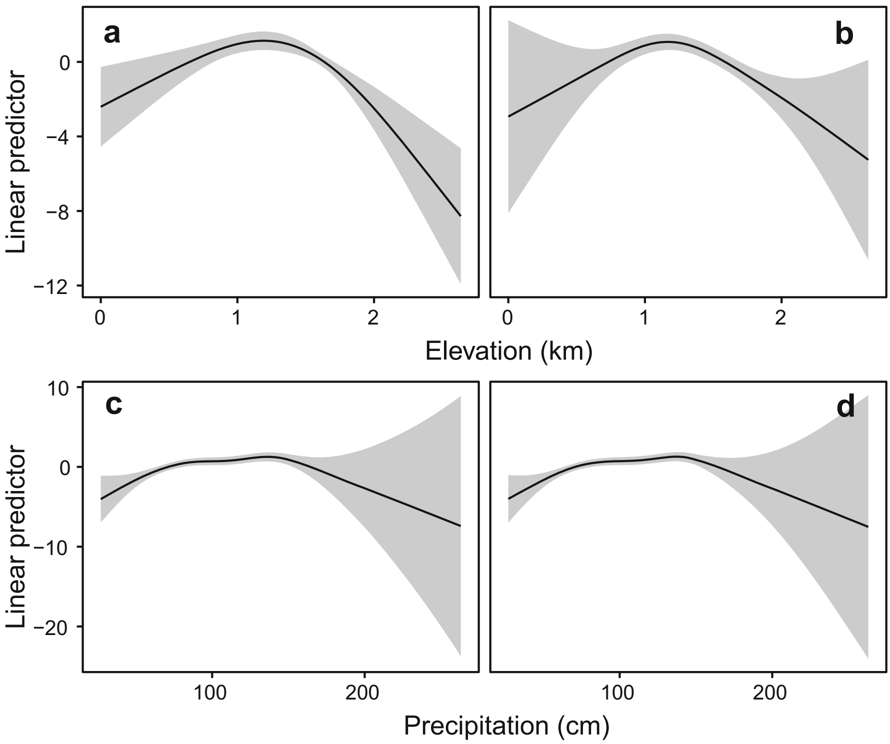

Fig. 7.11 Tuning a generalized additive model. Top panel contrasts (**a**, **b**) the number of knots used for modeling relationships of thrush occurrence with elevation (3 versus 6 knots). Bottom panel contrasts (**c**) Thin-plate spline versus (**d**) cubic spline (for automated knot selection) for precipitation

In this case, we find that there is some support for only having six knots in the smoother and the use of the tensor product. We can then map the model in a way similar to above:

```
> gam.map <- predict(layers, gam.vath.knot6, type = "response")
```

7.3.4.3 Regression Trees and Forests

Here, we focus on the application of Random Forest models using the `randomForest` package (Liaw and Wiener 2002). For Boosted Regression models, see the `gbm` package and the tutorial in Elith et al. (2008). The `randomForest` package can model both categorical (classification) and continuous (regression) response variables. We will implement a classification model. The default model function can be implemented as:

```
> library(randomForest)
> rf.vath <- randomForest(as.factor(pres) ~ canopy + elev +
  mesic1km + precip, data = all.cov, na.action = na.omit)
```

There are two parameters that are frequently adjusted for model tuning: `mtry` and `ntree`. `mtry` is the number of explanatory variables that are sampled for each tree, while `ntree` is the number of trees that are grown to produce the forest. We use the `tuneRF` function to determine the optimal values for `mtry`:

```
> rf.vath.tune <- tuneRF(y = as.factor(all.cov$pres), x =
  all.cov[,c(3:6)], stepFactor = 0.5, ntreeTry = 500)
```

Here we specify `ntreeTry` = 500, which is the default in the function. In general, it is thought that predictions are less sensitive to `ntree` than `mtry`. The `tuneRF` function adjusts `mtry` at different intervals (`stepFactor`), determining which value minimizes the predictive error (out-of-bag error). With this tuning, we update the model with `mtry=1` based on the out-of-bag error:

```
> rf.vath <- randomForest(as.factor(pres) ~ canopy + elev +
  mesic1km + precip, data = all.cov, mtry = 1, ntree = 500,
  na.action = na.omit)
```

We can then map the Random Forest prediction, similar to other models (Fig. 7.10).

```
> rf.map <- predict(layers, rf.vath, type = "prob", index = 2)
> plot(rf.map)
```

The primary difference here is that we specify `'index = 2'` because the `predict` function will make predictions for each class (there can be ≥2). In this case, it provides predictions for 0 and 1, with 1 being the second column from the predict object (thus, we ask for `index` = `2` to plot predictions).

7.3.4.4 Maximum Entropy

The use of maximum entropy for species distribution modeling relies on the Maxent program, which is a stand-alone Java software that is freely downloaded (http://biodiversityinformatics.amnh.org/open_source/maxent/). We can call this package in R via `dismo`. Note Phillips et al. (2017) also recently released the `maxnet` package for fitting Maxent models in R based on its relationship to the inhomogeneous point process model (see Sect. 7.2.5.5 for more). This package may be preferred in many cases because it does not require linking to the stand-alone Maxent program. We first focus on the use of dismo to call the Maxent software, because of its widespread use and useful interface for comparing across models, but briefly illustrate the use of maxnet in Sect. 7.2.5.5.

To call the stand-alone Maxent software from R, Java (https://www.oracle.com/java/index.html) must be installed on your computer. Note that if you run R on a 64 bit platform, you will need to make sure to install Java for 64 bit. Also, rJava (Urbanek 2017) will need to be installed and loaded in R. Once Maxent is downloaded, the maxent.jar file must be placed in the java folder of the dismo package. The location of this file can be found with the following:

```
> system.file("java", package = "dismo")
```

The maxent function calls Maxent to fit the model:

```
#default
> max.vath <- maxent(layers, p = vath.pres.xy)
```

The default maxent function takes presence-only points (vath.pres.xy) and generates 10,000 background points for comparison, extracting environmental data from these points and the presence points. We can manually provide background points instead, which can be useful to control the precise number and location of points used when comparing modeling techniques.

```
#default, but provide background points
> max.vath <- maxent(layers, p = vath.pres.xy, a = back.xy)
```

We can tune the Maxent model in several ways (Merow et al. 2013). Two common approaches are to: (1) adjust the regularization of the model; and (2) adjust the types of features considered (Fig. 7.12). We illustrate examples of both of these adjustments.

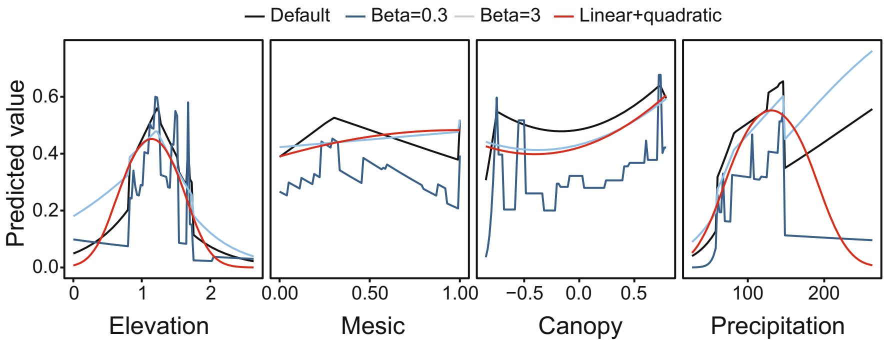

Fig. 7.12 Tuning a Maxent model. Shown are the default response curves, setting the regularization multiplier (β) to 0.3 and 3, and only considering linear and quadratic features

First, the regularization parameter can be changed manually. In this context, Maxent uses the lasso technique for regularization, such that coefficients that do not explain variation in presence locations are penalized and shrink toward zero. In this way, the default value for regulation is proportional to the number of presence locations and the variability in the environmental covariate at presence locations (Elith et al. 2011). The parameter, β, is a constant that is multiplied by the default regularization value. As β increases, a greater penalty is imposed. We can check this by adjusting beta and plotting changes in response curves (Fig. 7.12).

```
> maxent.beta3 <- maxent(layers, p = vath.pres.xy, a = back.xy,
  args = c("betamultiplier=3"))
```

In the above model, we specify a beta multiplier of 3 (the default setting is 1). Typically, this multiplier is altered to be > 1 because of concerns regarding potential overfitting of environmental relationships, but in Fig. 7.12 we illustrate setting the multiplier to be < 1 as well. We can also alter model complexity in terms of the features considered:

```
> maxent.features <- maxent(layers, p = vath.pres.xy, a =
  back.xy, args = c("noproduct", "nohinge", "nothreshold"))
```

In the above model, we tell Maxent to not use product (interactions), hinge, or threshold features. This reduces the model complexity to only consider linear and quadratic features, similar to a simple GLM. We can interpret the impacts of this tuning on partial relationships with the `dismo` package (see below for customizing partial plots) (Fig. 7.12):

```
> response(max.vath, expand = 0)
> response(maxent.beta3, expand = 0)
> response(maxent.features, expand = 0)
```

In the above response functions, we specify `expand = 0` to constrain the response plots only to the range of data considered. We can also evaluate the models with the `evaluate` function from the `dismo` package to get AUC statistics for each model. This function requires passing validation presence and absence points. Here, we use the validation samples (output not shown).

```
> evaluate(p = vath.val.pres, a = vath.val.abs, max.vath, layers)
> evaluate(p = vath.val.pres, a = vath.val.abs, maxent.beta3, layers)
> evaluate(p = vath.val.pres, a = vath.val.abs, maxent.features,
  layers)
```

This comparison suggests that each of these models are similar, in terms of AUC. See Sect. 7.2.7 *Model Evaluation* for a more detailed evaluation assessment. Finally, we can map the model (Fig. 7.10), similar to above :

```
> max.map <- predict(layers, max.vath)
```

Note that the prediction values for this Maxent model tend to be much higher than the GLM, GAM, and Random Forest model (Fig. 7.10). Why is that? Maxent provides different ways to plot and interpret the predictions. The default approach in this function is the “logistic” output, whereas the underlying Maxent model output is termed “raw” output. In the raw output, probabilities across the region sum to 1, such that the probability in any given location is very low and is essentially a probability density, sometimes referred to as relative occurrence rate (ROR; Merow et al. 2013). This can be requested in the `predict` function as:

```
> max.raw.map <- predict(layers, max.vath, args = "outputformat
  = raw")
```

The logistic output is a transformation of the raw output, aimed at providing probabilities that are more akin to probabilities of occurrence (Elith et al. 2011). In doing so, the average prediction for a location where a presence point occurrence with the logistic output approaches 0.5. Another alternative to the logistic and raw outputs is the cumulative log-log (cloglog) output (Fithian et al. 2015), which is better rooted in probability theory and is now the default output in the stand-alone Maxent software (Phillips et al. 2017). These different response outputs should not change the rank suitabilities from models, but they will change the absolute values such that care should be taken when implementing and interpreting output.

7.3.4.5 Point Process Models

Finally, we note that most of the above models can be recast formally as inhomogeneous point process (IPP) models. There are several benefits for doing so, because this perspective provides a means to better understand the number of background points needed, understand the role of spatial dependence, and interpret goodness-of-fit and related model diagnostics (Fithian and Hastie 2013; Phillips et al. 2017).

To implement the above models as point process models, Renner et al. (2015) suggested that many more background points should be considered because they are interpreted as “quadrature” points used for approximating an integral in the point process function that describes the background environment. Warton and Shepherd (2010) argued that it is natural to do so by creating a grid of background points (rather than random point generation), which could be created with the `sampleRegular` function in the `raster` package. With these points, point process models can be fit with a variety of packages. A simple updating of the above GLMs and GAMs with a point process formulation would be (Renner et al. 2015):

```
> glm.ppm <- glm(pres ~ canopy + elev + I(elev^2) + mesic1km +
  precip, family = binomial(link=logit), weights = 1000^(1-pres),
  data = all.cov)
```

```
> gam.ppm <- gam(pres ~ s(canopy) + s(elev) + s(mesic1km) +
  s(precip), family = binomial(link = logit), weights = 1000^(1-
  pres), data = all.cov)
```

We use weighted regressions in the above models to approximate the inhomogeneous point process, where we provide arbitrarily large weights to the background points. Also note that when implementing this model, we should include a larger number of background points than what is shown here, potentially sampled in a regular grid. The number of background points can be formally determined in this context by altering the number of background points until the likelihood of the model stabilizes (Renner and Warton 2013; Renner et al. 2015).

We can also fit a Maxent model with a point process formulation using the maxent function in the dismo package:

```
> maxent.ppm <- maxent(layers, p = vath.pres.xy, a = back.xy,
  args = c("noremoveduplicates"))
```

The key difference in the above model is that in the maxent function we specify to not remove duplicate records (multiple presence locations within a cell on the map). If we did not pass our own background points, we would also need to add "noaddsamplestobackground" and increase the number of background points generated (e.g., "maximumbackground = 50000" for 50,000 points).

Finally, the maxnet package can be used as well, which uses the glmnet package (Friedman et al. 2010) to fit a Maxent-formulated IPP (based on the idea of "infinitely weighted logistic regression") (Fithian and Hastie 2013) that uses the same regularization and features that the stand-alone Maxent package provides. In this case, maxnet requires a different data format than the maxent function in dismo, where we provide a vector of presence and background locations and a data frame or matrix of covariates at those locations:

```
> library(maxnet)
> library(glmnet)
> max.cov <- all.cov[,c("canopy", "elev", "mesic1km", "precip")]
> maxnet.ppm <- maxnet(all.cov$pres, max.cov)
```

In this function, features can be requested and the regularization constant adjusted in the following way:

```
> maxnet.beta3.linquad <- maxnet(all.cov$pres, max.cov, regmult
  = 3, maxnet.formula(all.cov$pres, max.cov, classes = "lq"))
```

The classes statement provides a means to select features for the model, with all features being 'lqhpt' (linear, quadratic, hinge, product, threshold). We do not focus on these model IPP formulations below, but the interested reader should see Renner et al. (2015).

7.3.5 Interpreting Environmental Relationships

Each of these distribution modeling algorithms uses some sort of function regarding species distribution and environmental factors. A challenge is to interpret these functions in a meaningful way across models.

A common approach to do so is through use of partial response plots (or "partial plots"). In these plots, we vary one environmental covariate across the range of observed variation while setting all other environmental covariates to a constant, typically their mean or median. We then make predictions on this new data set to interpret how the models are relating species occurrence to environmental factors. Note that this approach will not adequately illuminate potential interactions between variables if they are considered in models (e.g., through the use of tensor products in GAMs or in Random Forest models). However, it can still be useful for interpreting patterns that underlie the predictions for each algorithm. Elith et al. (2005) generalized this idea to model algorithms that only make predictions on raster grids with what they term the "evaluation strip," or the addition of data to a raster grid that serves a similar purpose as making predictions to new data with partial response plots.

There are some packages that provide functions for calculating partial plots (e.g., the `response` function used above in the `dismo` package), and some of the wrapper packages, such as `biomod2` provide general functions in this way. Here, we illustrate how users can manually accomplish this task, which provides a means to alter graphics or subtle aspects of predictions (e.g., adding uncertainty in predictions). The following code focuses on creating partial plots for the elevation covariate, but see Fig. 7.13 for plots of all covariates. We first generate a new data set (`elev.partial.data`) for predictions:

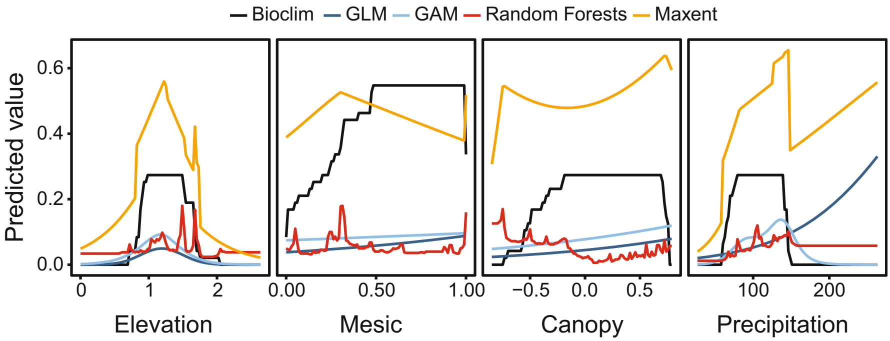

Fig. 7.13 Partial plots from models for elevation (km), proportion of mesic forest within 1km, canopy cover (relative units based on a PCA), and precipitation (cm). For each covariate, all other covariates were set to their median value

```
> canopy.median <- median(back.cov$canopy)
> precip.median <- median(back.cov$precip)
> mesic1km.median <- median(back.cov$ mesic1km)

> elev.range <- seq(min(back.cov$elev),
  max(back.cov$Elev), length = 100)
```

We put the covariates into a data frame and use the `expand.grid` function to expand the data for all possible combinations:

```
> elev.partial.data <- data.frame(expand.grid(Elev = elev.range,
  Canopy = canopy.median, precip = precip.median, mesic1km =
  mesic1km.median))
```

We then make predictions from each model onto this new data set:

```
> bio.pred.elev <- predict(bioclim.vath, elev.partial.data)
> glm.pred.elev <- predict(glm.vath, elev.partial.data, type =
      "response")
> gam.pred.elev <- predict(gam.vath, elev.partial.data, type =
      "response")
> rf.pred.elev <- predict(rf.vath, elev.partial.data, type =
  "prob")
> rf.pred.elev <- rf.pred.elev[,2]
> max.pred.elev <- predict(max.vath, elev.partial.data)
```

Finally, we can use the `plot` function or `ggplot2` (Wickham 2009) to create a partial prediction plot. Here we show the use of `plot` to illustrate plotting the Bioclim, GLM, and Random Forest predictions.

```
#create data frame
> part.elev.df <- data.frame(elevation = elev.range, bioclim =
  bio.pred.elev, glm = glm.pred.elev, gam = gam.pred.elev, rf =
  rf.pred.elev, max = max.pred.elev)

#plot
> plot(part.elev.df$elevation, part.elev.df $bioclim, type = 'l')
> lines(part.elev.df$elevation, part.elev.df$glm, type = 'l',
  col = "red")
> lines(part.elev.df$elevation, part.elev.df$rf, type = 'l', col
  = "blue")
```

These partial plots illustrate the widely divergent environmental functions identified across algorithms in modeling thrush occurrence (Fig. 7.13). Overall, the partial responses for Random Forests are highly complex and non-linear, while the partial responses for the other algorithms are smoother and less complex. Note that the absolute predictions vary as well. This pattern occurs between the modeling algorithms because of the different currencies that they are modeling. Bioclim is modeling similarity and the predictions from Maxent are based on the logistic output, which tends to make the average prediction for presence locations approximately 0.5 (Elith et al. 2011). In contrast, the GLM, GAM and Random Forests are discriminating presence points versus background points, such that as we increase the number of background points, the probabilities on the *y*-axis will decrease (because increasing the background points decreases the intercept term value in the model). For instance, if we generated the same number of background points as presence points, the intercept on these latter models would generate predictions on these partial plots with means close to 0.5.

7.3.6 Model Evaluation

The above models can be evaluated in a variety of ways and there several packages for model evaluation. The `dismo` package includes the `evaluation` function, but here we use the `PresenceAbsence` package (Freeman and Moisen 2008), which includes a more comprehensive set of evaluation metrics. To use the `PresenceAbsence` package, we create a data frame that includes (in the following order): (1) site IDs for the validation (evaluation) data; (2) the observed responses in the validation data; and (3) model predictions for those locations. This data frame can have predictions from *N* models, where columns for predictions are 3 to *N*+3. We first illustrate model evaluation based on the prospective sampling dataset from 3 to 4 years later in time, and then illustrate how model evaluation can be accomplished with *K*-fold validation (Boyce et al. 2002), a common approach to model evaluation.

For the prospective sampling validation data set, we simply take each of the above models and make predictions for the new locations:

```
> val.cov.pred <- val.cov[,cbind("canopy", "elev", "mesic1km",
  "precip")]
> bio.val <- predict(bioclim.vath, val.cov.pred)
> glm.val <- predict(glm.vath, val.cov.pred, type = "response")
> gam.val <- predict(gam.vath, val.cov.pred, type = "response")
> rf.val <- predict(rf.vath, val.cov.pred, type = "prob")
> rf.val <- rf.val[,2]
> max.val <- predict(max.vath, val.cov.pred)
```

With these predictions, we then create a data frame that is formatted for the `PresenceAbsence` package and we will create a data frame for storing the model evaluation results.

```
> val.data <- data.frame(siteID = 1:nrow(vath.val), obs =
  vath.val$VATH, bio = bio.val, glm = glm.val, gam = gam.val, rf =
  rf.val, max = max.val)
```

```
> summary.eval <- data.frame(matrix(nrow = 0,ncol = 9))
> names(summary.eval)<-c("model", "auc", "corr", "ll",
  "threshold", "sens", "spec", "tss", "kappa")
```

For model evaluation, we will calculate three continuous metrics: AUC, the biserial correlation coefficient, and the cross-validated log-likelihood (Lawson et al. 2014). We will also calculate four binary metrics taken from the confusion matrix: sensitivity, specificity, kappa, and the true skill statistic. The `PresenceAbsence` package can determine thresholds based on a variety of criteria, such as prevalence in the test or training data, maximizing kappa or maximizing the sum of specificity and sensitivity (see `?optimal.thresholds`). Here, we focus on using a threshold that maximizes the sum of specificity and sensitivity (`opt.methods=3` in the `optimal.thresholds` function), which was recommended by Liu et al. (2013). In the following `for` loop, we calculate each of these metrics for each model and populate our summary data frame with the output. We first load the `PresenceAbsence` package and detach `glmnet`, because the latter package also includes a function for calculating AUC.

```
> library(PresenceAbsence)
> detach("package:glmnet")
> nmodels <- ncol(val.data)−2
> for(i in 1:nmodels){
  auc.i <- auc(val.data, which.model = i)
  kappa.opt <- optimal.thresholds(val.data, which.model = i,
   opt.methods = 3)
  sens.i <- sensitivity(cmx(val.data, which.model = i, threshold
= kappa.opt[[2]]))
  spec.i <- specificity(cmx(val.data, which.model = i, threshold
= kappa.opt[[2]]))
  tss.i <- sens.i$sensitivity + spec.i$specificity − 1
  kappa.i <- Kappa(cmx(val.data, which.model = i, threshold =
   kappa.opt[[2]]))
  corr.i <- cor.test(val.data[,2], val.data[,i + 2])$estimate
  ll.i <- sum(log(val.data[,i + 2] * val.data[,2] + (1 −
  val.data[,i + 2]) * (1 − val.data[,2])))
  ll.i <- ifelse(ll.i == "−Inf", sum(log(val.data[,i + 2] +
  0.01) * val.data[,2] + log((1 − val.data[,i + 2])) * (1 −
  val.data[,2])), ll.i)
  summary.i <- c(i, auc.i$AUC, corr.i, ll.i, kappa.opt[[2]],
   sens.i$sensitivity, spec.i$specificity, tss.i, kappa.i[[1]])
  summary.eval <- rbind(summary.eval, summary.i)
}
```

Note that in the above code, we add a small constant to the log-likelihood calculation because the log(0) is undefined (e.g., when the predicted value is 0, as can be the case in the Bioclim model). Based on these summary statistics, it is clear that none of these models appear to predict well to the prospective sampling data set (Table 7.4), despite the fact that these models had clear environmental relationships (see, e.g., `summary(glm.vath)` and `summary(gam.vath)`). This result

Table 7.4 Evaluation of modeling algorithms based on external validation (presence–absence data collected 3–4 years later)

Model	AUC	LL_{cv}	TSS	Kappa
Bioclim	0.586	−685	0.136	0.027
GLM	0.673	−519	0.287	0.106
GAM	0.651	−528	0.237	0.092
Random Forests	0.625	−607	0.182	0.039
Maxent	0.669	−971	0.259	0.164

illustrates the potential challenge of generating accurate species distribution models that can predict accurately over time (Eskildsen et al. 2013; Vallecillo et al. 2009). In this case, the Bioclim and Random Forests models tended to predict the worst of the models considered based on model discrimination, whereas the Maxent model predicts poorly using the logistic output based on the cross-validated log-likelihood, a model calibration metric. We include cross-validated log-likelihoods because they are useful calibration metrics (Lawson et al. 2014); however, they are most properly applied to models trained with presence–absence data rather than presence-only data.

We can also evaluate models with calibration plots. Calibration plots can be easily generated with the `PresenceAbsence` package. For the above models, we use the `calibration.plot` function. An example for the Maxent model is:

```
> calibration.plot(val.data, which.model = 5, N.bins = 5, xlab =
  "Predicted", ylab = "Observed", main = "maxent")
```

Note that this function requires the user to define the number of bins that will be used to pool binary data.

A more common approach is to use *K*-fold validation (Boyce et al. 2002). In that approach we subset the training data into subsets, or folds. We then fit models holding out one fold while using $K-1$ folds for model training. This is then repeated *K* times. The above evaluation code can be readily applied each fold and then we summarize across folds. We consider fivefolds and apply this approach to the presence–background data used for model training. The `dismo` package has a function `kfold` that will create a vector of *k* groups based on random allocation to groups, with the constraint that each group is of equal size.

In this *K*-fold case, we are using presence–background data for model evaluation. In general, using such data for model evaluation is limited because no absence data are available for evaluation. In such situations, it is often recommended to use evaluation metrics that only make use of information on presence locations (Guisan et al. 2017). Here, we use the Boyce index, a common metric for evaluating presence-only predictions that does not rely on absence data (Boyce et al. 2002), which can be calculated with the `ecospat` package (Broennimann et al. 2018). We also calculate the same metrics as above for illustrative purposes. In practice, the Boyce index and other metrics aimed specifically for evaluation with presence-only data should be emphasized (Engler et al. 2004; Hirzel et al. 2004). See Guisan et al. (2017) for more information.

```
#number of k-folds considered
> folds <- 5

#create k-folds
> kfold_pres <- kfold(pres.cov, k = folds)
> kfold_back <- kfold(back.cov, k = folds)
```

Above we apply the `kfold` function separately to the presence and background data. This ensures that each fold will contain the same number of presence points. Then we can apply a `for` loop or something similar to go through each fold. We do not provide the entire for loop here but illustrate how data can be subset for each fold, *k*.

```
#partition data based on each k-fold
 > kfolds <- 1
 > val.pres.k <- pres.cov[kfold_pres == kfolds, ]
 > val.back.k <- back.cov[kfold_back == kfolds, ]
 > val.k <- rbind(val.pres.k, val.back.k)

 > train.pres.k <- pres.cov[kfold_pres != kfolds, ]
 > train.back.k <- back.cov[kfold_back != kfolds, ]
 > train.k <- rbind(train.pres.k, train.back.k)
```

We apply each of these new data sets (either `train.k`, or each component of the training data set, depending on the model algorithm) to the model algorithms of interest described above and make predictions onto the validation data (`val.k`). With this data format, the Boyce index can be calculated for a given model `i` within the `for` loop mentioned above regarding external validation as:

```
> library(ecospat)
> boyce.i <- ecospat.boyce(fit = val.data[,i + 2], obs =
  val.data[1:nrow(val.pres.k),i + 2], res = 100, PEplot = F)
```

Note that `Biomod2` and `sdm` have built-in cross-validation functions; however, here we illustrate how to accomplish *K*-fold validation manually, which allows users to customize how *K*-fold validation is accomplished (see also `ecospat` for functions regarding cross-validation). For example, there has been recent criticism regarding how folds are delineated (Hijmans 2012). In the code above, we randomly select points to folds, yet these points are not likely spatially independent. An alternative is to use "block" *K*-fold validation, where spatial blocks are randomly selected, rather than sample points (Wenger and Olden 2012). In this case, we might randomly select transects or watersheds as blocks for validation purposes. This would be straightforward to accomplish above by sampling transects in lieu of points.

Table 7.5 Evaluating modeling algorithms based on *K*-fold validation

Model	Boyce	AUC	TSS	Kappa
Bioclim	0.525	0.737	0.440	0.080
GLM	0.737	0.781	0.473	0.156
GAM	0.798	0.802	0.462	0.135
Random Forests	0.791	0.839	0.572	0.211
Maxent	0.851	0.803	0.500	0.154

Based on *K*-fold validation, we get a different perspective on the utility of these models (Table 7.5), where summary statistics tend to be higher than with prospective sampling. We also find that the more complex models tend to be favored more, with the Random Forest model tending to predict relatively better when using *K*-fold validation.

7.3.7 Combining Models: Ensembles

With predicted maps, it is straightforward to create model ensembles. A common approach is to make a weighted average of predictions from models based on AUC for each model or some other model evaluation metric (Marmion et al. 2009). We emphasize, however, that because different algorithms model different currencies, we suggest that averaging of predictions should only be made for models that are modeling the same currency. Here, we show how we can create an ensemble based on the GLM and GAM model (Fig. 7.14), which are predicting the same response quantity (and thus on similar currencies; unlike Bioclim and Maxent), but differ in their environmental functions being considered.

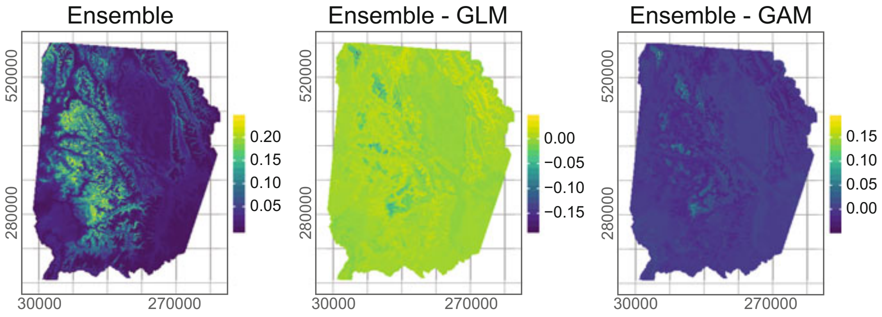

Fig. 7.14 (**a**) An ensemble from the GLM and GAM using a weighted mean based on AUC scores taken from *K*-fold validation, and the difference in ensemble predictions and predictions from (**b**) the GLM, and (**c**) the GAM

```
> models <- stack(glm.map, gam.map)
> names(models) <- c("glm", "gam")

#weighted average based on AUC from prospective sampling
> AUC.glm <- summary.eval[summary.eval$model == "glm", "auc"]
> AUC.gam <- summary.eval[summary.eval$model == "gam", "auc"]
> auc.weight <- c(AUC.glm, AUC.gam)

> ensemble.auc <- weighted.mean(models, auc.weight)
> plot(ensemble.auc)
```

Other approaches to ensemble modeling can include truncating predictions to binary information of predicted presence/absence and then summarizing this information in a variety of ways. Such truncation might be preferred when combining very different modeling techniques. For instance, with that information, models could be integrated by: (1) quantifying a bounded box of predicted occurrence, or the region where at least one algorithm predicts occurrence; or (2) mapping the frequency of predicted occurrence from different model algorithms (Araújo and New 2007) (Fig. 7.8).

7.4 Next Steps and Advanced Issues

7.4.1 Incorporating Dispersal

A common criticism for distribution models is that they typically ignore dispersal-related limitations (Barve et al. 2011). Some approaches simply apply constraints to the mapping process (Cardador et al. 2014), some model colonization processes with time series data (see Chap. 10) (Bled et al. 2013; Yackulic et al. 2015), while others link distribution models with simulations of the dispersal process (Smolik et al. 2010). In general, there is a great need to incorporate movement into the prediction of species distributions and this is an active area of development (Miller and Holloway 2015; Boulangeat et al. 2012). The `MigClim` package provides some functionality for incorporating dispersal constraints into distribution modeling (Engler et al. 2012).

7.4.2 Integrating Multiple Data Sources

Given the limitations of presence-only data and yet the wide availability of such data, it is tempting to integrate presence-only data with other data that suffer fewer biases. Recent modeling advances aim to unite opportunistic presence-only data with presence–absence, occupancy or abundance data to make more reliable predictions by using multiple sources of data simultaneously in model development, termed integrated species distribution models (Fithian et al. 2015; Koshkina et al. 2017; Pacifici et al. 2017). Such integration can help to minimize bias as well as providing a means to incorporate species prevalence, which is helpful for making predictions of the probability of occurrence. These modeling efforts have been shown to reduce potential bias and increase predictive accuracy of models (Dorazio 2014; Fithian et al. 2015; Fletcher et al. 2016).

7.4.3 Dynamic Models

When time series of location data occur, we may model dynamics of distributions. In this approach, often the focus is on understanding local extinction-colonization dynamics (MacKenzie et al. 2003; Yackulic et al. 2015). In this framework, species distribution (e.g., occupancy) over time is a derived parameter from estimated colonization-extinction dynamics. Benefits of modeling dynamics include that it provides a means to better understand the role of different limiting factors on species distribution (e.g., dispersal limitation) (Broms et al. 2016), whether occurrence at locations reflect underlying habitat quality (Pulliam 2000), it can help identify if species distributions tend to be at equilibrium (a prominent assumption when projecting distributions) (Yackulic et al. 2015), and it can allow predictions of range dynamics (Guillera-Arroita 2017). We will address these dynamical models in Chap. 10.

7.4.4 Multi-species Models

There is also increasing interest in modeling the distribution of multiple species simultaneously (Ferrier and Guisan 2006). This can be done in a variety of ways, and it typically focuses on species co-occurrence (Dorazio et al. 2006; Ferrier et al. 2007; Ovaskainen et al. 2010; Araújo et al. 2011). Simultaneously modeling multiple species can be advantageous for several reasons. First, it can provide insight into potential species interactions and how those change over space or time. Second, one species might be a good predictor of another species, not necessarily due to interactions but because it is an indirect indicator of environmental conditions. We will address some of these techniques in Chap. 11.

7.4.5 Sampling Error and Distribution Models

Throughout this chapter, we have ignored the problem of sampling error, such as imperfect detection of species, to focus more simply on the issues variation in model algorithms and model evaluation. However, observation errors are common in data sets and these errors frequently need to be accounted for to obtain reliable estimates of environmental relationships. Several models exist for accounting for imperfect detection, both in terms of false positive and false negative errors (Miller et al. 2011; Guillera-Arroita et al. 2017). False negative errors are more common, where a species or individual occurs in an area but we fail to detect it. Several investigations suggest that accounting for false negative errors can improve the predictive performance of distribution models (Rota et al. 2011; Lahoz-Monfort et al. 2014; Guillera-Arroita 2017). One major challenge in the interpretation, however, is that these models predict occupancy across a geographic region, and yet evaluation data are often detection-non detection (typically true occupancy data are not available for evaluating models). Distribution models that account for imperfect detection can be fit with a variety of R packages, including the `unmarked` (Fiske and Chandler 2011), `hSDM` (Vielledent et al. 2014), and `stocc` (Johnson 2015) packages.

7.5 Conclusions

Understanding, predicting, and projecting species distributions provides a means to answer major questions in ecology and can deliver decision support for many conservation problems (Gill et al. 2001; Norris 2004; Wiens et al. 2010; Guisan et al. 2013). The use of species distribution models in ecology, evolution, and conservation has a long tradition (Rotenberry and Wiens 1980; Austin 1987; Donovan et al. 1987), and yet it has exploded over the past 15 years with new advances in modeling algorithms and newly available data sources regarding species locations and geo-spatial data of environmental factors (Graham et al. 2004; Dickinson et al. 2010; Fick and Hijmans 2017).

Many of the species distribution modeling techniques currently being used can be described as inhomogeneous point process models. This realization has several consequences for the implementation and interpretation of species distribution models (Renner and Warton 2013; Renner et al. 2015). We recommend that this framework be generally used to guide correlative species distribution modeling.

Our example illustrates that reliably applying and evaluating species distribution models can be challenging. Extrapolating predictions beyond the environmental conditions used for model building, an issue that commonly occurs when projecting the effects of climate change (Thomas et al. 2004), can be difficult because little information exists on such relationships. Evaluating models with commonly used techniques, such as *K*-fold validation, can sometimes provide a false sense of model performance (Wenger and Olden 2012) and suggest that more complex models are valuable when in fact simpler models may be sufficient for reliable predictions in space and time (cf. Tables 7.4 and 7.5).

Despite this increased use distribution models, these models still have limitations and there use and application should be done with care. Greater focus on mechanistic modeling and leveraging information on why species distribution varies over space and time may further advance our understanding of species distribution and our ability to predict changes in distribution with ongoing environmental change.

References

Aarts G, Fieberg J, Matthiopoulos J (2012) Comparative interpretation of count, presence-absence and point methods for species distribution models. Methods Ecol Evol 3(1):177–187. https://doi.org/10.1111/j.2041-210X.2011.00141.x

Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). J Appl Ecol 43(6):1223–1232

Araujo MB, Guisan A (2006) Five (or so) challenges for species distribution modelling. J Biogeogr 33(10):1677–1688

Araújo MB, New M (2007) Ensemble forecasting of species distributions. Trends Ecol Evol 22 (1):42–47. https://doi.org/10.1016/j.tree.2006.09.010

Araújo MB, Peterson AT (2012) Uses and misuses of bioclimatic envelope modeling. Ecology 93 (7):1527–1539

Araújo MB, Rozenfeld A, Rahbek C, Marquet PA (2011) Using species co-occurrence networks to assess the impacts of climate change. Ecography 34(6):897–908. https://doi.org/10.1111/j.1600-0587.2011.06919.x

Austin MP (1987) Models for the analysis of species response to environmental gradients. Vegetatio 69(1–3):35–45. https://doi.org/10.1007/bf00038685

Austin MP (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecol Model 157(2–3):101–118

Austin M (2007) Species distribution models and ecological theory: a critical assessment and some possible new approaches. Ecol Model 200(1–2):1–19

Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting pseudo-absences for species distribution models: how, where and how many? Methods Ecol Evol 3(2):327–338. https://doi.org/10.1111/j.2041-210X.2011.00172.x

Barton K (2018) MuMIn: multi-model inference. R package version 1.40.4

Barve N, Barve V, Jimenez-Valverde A, Lira-Noriega A, Maher SP, Peterson AT, Soberon J, Villalobos F (2011) The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecol Model 222(11):1810–1819. https://doi.org/10.1016/j.ecolmodel.2011.02.011

Berman M, Turner TR (1992) Approximating point process likelihoods with GLIM. J R Stat Soc C Appl Stat 41(1):31–38

Betts MG, Phalan B, Frey SJK, Rousseau JS, Yang ZQ (2018) Old-growth forests buffer climate-sensitive bird populations from warming. Divers Distrib 24(4):439–447. https://doi.org/10.1111/ddi.12688

Bled F, Nichols JD, Altwegg R (2013) Dynamic occupancy models for analyzing species' range dynamics across large geographic scales. Ecol Evol 3(15):4896–4909. https://doi.org/10.1002/ece3.858

Blonder B, Lamanna C, Violle C, Enquist BJ (2014) The n-dimensional hypervolume. Glob Ecol Biogeogr 23(5):595–609. https://doi.org/10.1111/geb.12146

Boulangeat I, Gravel D, Thuiller W (2012) Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. Ecol Lett 15(6):584–593. https://doi.org/10.1111/j.1461-0248.2012.01772.x

Boyce MS, Vernier PR, Nielsen SE, Schmiegelow FKA (2002) Evaluating resource selection functions. Ecol Model 157(2–3):281–300. https://doi.org/10.1016/s0304-3800(02)00200-4

Brand LA, George TL (2001) Response of passerine birds to forest edge in coast redwood forest fragments. Auk 118(3):678–686. https://doi.org/10.1642/0004-8038(2001)118[0678:Ropbtf]2.0.Co;2

Breiman L (2001) Random forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/a:1010933404324

Breiman L, Friedman J, Stone CJ, Olshen RA (1984) Classification and regression trees. Chapman and Hall/CRC, Boca Raton, FL

Brewer CK, Berglund D, Barber JA, Bush R (2004) Northern region vegetation mapping project summary report and spatial datasets, version 42. Northern Region USFS

Broennimann O, Treier UA, Muller-Scharer H, Thuiller W, Peterson AT, Guisan A (2007) Evidence of climatic niche shift during biological invasion. Ecol Lett 10(8):701–709. https://doi.org/10.1111/j.1461-0248.2007.01060.x

Broennimann O, Di Cola V, Guisan A (2018) ecospat: spatial ecology miscellaneous methods. R package verision 3.0

Broms KM, Hooten MB, Johnson DS, Altwegg R, Conquest LL (2016) Dynamic occupancy models for explicit colonization processes. Ecology 97(1):194–204. https://doi.org/10.1890/15-0416.1

Brotons L, Thuiller W, Araujo MB, Hirzel AH (2004) Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. Ecography 27(4):437–448. https://doi.org/10.1111/j.0906-7590.2004.03764.x

Buckley LB (2008) Linking traits to energetics and population dynamics to predict lizard ranges in changing environments. Am Nat 171(1):E1–E19. https://doi.org/10.1086/523949

Buckley LB, Kingsolver JG (2012) Functional and phylogenetic approaches to forecasting species' responses to climate change. Annu Rev Ecol Evol Syst 43:205–226. https://doi.org/10.1146/annurev-ecolsys-110411-160516

Buckley LB, Urban MC, Angilletta MJ, Crozier LG, Rissler LJ, Sears MW (2010) Can mechanism inform species' distribution models? Ecol Lett 13(8):1041–1054. https://doi.org/10.1111/j.1461-0248.2010.01479.x

Burnham KP, Anderson DR (1998) Model selection and inference: a practical information-theoretic approach. Springer, New York

Busby JR (1991) BIOCLIM: a bioclimate analysis and prediction system. In: Margules CR, Austin MP (eds) Nature conservation: cost effective biological surveys and data analysis. CSIRO, Canberra, Australia, pp 64–68

Cardador L, Sarda-Palomera F, Carrete M, Manosa S (2014) Incorporating spatial constraints in different periods of the annual cycle improves species distribution model performance for a highly mobile bird species. Divers Distrib 20(5):515–528. https://doi.org/10.1111/ddi.12156

Carpenter G, Gillison AN, Winter J (1993) DOMAIN—a flexible modeling procedure for mapping potential distributions of plants and animals. Biodivers Conserv 2(6):667–680. https://doi.org/10.1007/bf00051966

Chase JM, Leibold MA (2003) Ecological niches: linking classical and contemporary approaches. University of Chicago Press

Colwell RK, Rangel TF (2009) Hutchinson's duality: the once and future niche. Proc Natl Acad Sci U S A 106:19651–19658. https://doi.org/10.1073/pnas.0901650106

Cushman SA, McGarigal K (2004) Patterns in the species-environment relationship depend on both scale and choice of response variables. Oikos 105(1):117–124

Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT (2007) Random forests for classification in ecology. Ecology 88(11):2783–2792. https://doi.org/10.1890/07-0539.1

De'ath G, Fabricius KE (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81(11):3178–3192. https://doi.org/10.1890/0012-9658(2000)081[3178:Cartap]2.0.Co;2

Di Cola V, Broennimann O, Petitpierre B, Breiner FT, D'Amen M, Randin C, Engler R, Pottier J, Pio D, Dubuis A, Pellissier L, Mateo RG, Hordijk W, Salamin N, Guisan A (2017) ecospat: an R package to support spatial analyses and modeling of species niches and distributions. Ecography 40(6):774–787. https://doi.org/10.1111/ecog.02671

Dickinson JL, Zuckerberg B, Bonter DN (2010) Citizen science as an ecological research tool: challenges and benefits. Annu Rev Ecol Evol Syst 41:149–172. https://doi.org/10.1146/annurev-ecolsys-102209-144636

Donovan ML, Rabe DL, Olson CE (1987) Use of geographic information-systems to develop habitat suitability models. Wildl Soc Bull 15(4):574–579

Dorazio RM (2014) Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. Glob Ecol Biogeogr 23(12):1472–1484. https://doi.org/10.1111/geb.12216

Dorazio RM, Royle JA, Soderstrom B, Glimskar A (2006) Estimating species richness and accumulation by modeling species occurrence and detectability. Ecology 87(4):842–854. https://doi.org/10.1890/0012-9658(2006)87[842:esraab]2.0.co;2

Dormann CF, Schymanski SJ, Cabral J, Chuine I, Graham C, Hartig F, Kearney M, Morin X, Roemermann C, Schroeder B, Singer A (2012) Correlation and process in species distribution models: bridging a dichotomy. J Biogeogr 39(12):2119–2131. https://doi.org/10.1111/j.1365-2699.2011.02659.x

Elith J, Leathwick JR (2009) Species distribution models: ecological explanation and prediction across space and time. Annu Rev Ecol Evol Syst 40:677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Elith J, Ferrier S, Huettmann F, Leathwick J (2005) The evaluation strip: a new and robust method for plotting predicted responses from species distribution models. Ecol Model 186(3):280–289. https://doi.org/10.1016/j.ecolmodel.2004.12.007

Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberon J, Williams S, Wisz MS, Zimmermann NE (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29 (2):129–151

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. J Anim Ecol 77 (4):802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Elith J, Kearney M, Phillips S (2010) The art of modelling range-shifting species. Methods Ecol Evol 1(4):330–342. https://doi.org/10.1111/j.2041-210X.2010.00036.x

Elith J, Phillips SJ, Hastie T, Dudik M, Chee YE, Yates CJ (2011) A statistical explanation of MaxEnt for ecologists. Divers Distrib 17(1):43–57. https://doi.org/10.1111/j.1472-4642.2010.00725.x

Elton C (1927) Animal ecology. Sedgwick and Jackson, London

Engler R, Guisan A, Rechsteiner L (2004) An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. J Appl Ecol 41 (2):263–274

Engler R, Hordijk W, Guisan A (2012) The MIGCLIM R package—seamless integration of dispersal constraints into projections of species distribution models. Ecography 35(10):872–878. https://doi.org/10.1111/j.1600-0587.2012.07608.x

Eskildsen A, le Roux PC, Heikkinen RK, Hoye TT, Kissling WD, Poyry J, Wisz MS, Luoto M (2013) Testing species distribution models across space and time: high latitude butterflies and recent warming. Glob Ecol Biogeogr 22(12):1293–1303. https://doi.org/10.1111/geb.12078

Evans JM, Fletcher RJ Jr, Alavalapati J (2010) Using species distribution models to identify suitable areas for biofuel feedstock production. Glob Change Biol Bioenergy 2(2):63–78. https://doi.org/10.1111/j.1757-1707.2010.01040.x

Feeley KJ, Silman MR (2010) Land-use and climate change effects on population size and extinction risk of Andean plants. Glob Chang Biol 16(12):3215–3222. https://doi.org/10.1111/j.1365-2486.2010.02197.x

Ferrier S, Guisan A (2006) Spatial modelling of biodiversity at the community level. J Appl Ecol 43 (3):393–404. https://doi.org/10.1111/j.1365-2664.2006.01149.x

Ferrier S, Manion G, Elith J, Richardson K (2007) Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. Divers Distrib 13(3):252–264. https://doi.org/10.1111/j.1472-4642.2007.00341.x

Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int J Climatol 37(12):4302–4315. https://doi.org/10.1002/joc.5086

Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ Conserv 24(1):38–49

Fiske IJ, Chandler RB (2011) Unmarked: an R package for fitting hierarchical models of wildlife occurrence and abundance. J Stat Softw 43(10):1–23

Fithian W, Hastie T (2013) Finite-sample equivalence in statistical models for presence-only data. Ann Appl Stat 7(4):1917–1939. https://doi.org/10.1214/13-aoas667

Fithian W, Elith J, Hastie T, Keith DA (2015) Bias correction in species distribution models: pooling survey and collection data for multiple species. Methods Ecol Evol 6(4):424–438. https://doi.org/10.1111/2041-210x.12242

Fletcher RJ Jr, Hutto RL (2008) Partitioning the multi-scale effects of human activity on the occurrence of riparian forest birds. Landsc Ecol 23:727–739

Fletcher RJ, McCleery RA, Greene DU, Tye CA (2016) Integrated models that unite local and regional data reveal larger-scale environmental relationships and improve predictions of species distributions. Landsc Ecol 31(6):1369–1382. https://doi.org/10.1007/s10980-015-0327-9

Franklin J (2009) Mapping species distributions: spatial inference and prediction. Cambridge University Press, Cambridge, UK

Freeman EA, Moisen G (2008) PresenceAbsence: an R package for presence absence analysis. J Stat Softw 23(11):1–31

Fretwell SD, Lucas HL Jr (1970) On territorial behavior and other factors influencing habitat distribution in birds. I. Theoretical development. Acta Biotheor 19:16–36

Friedman JH (2002) Stochastic gradient boosting. Comput Stat Data Anal 38(4):367–378. https://doi.org/10.1016/s0167-9473(01)00065-2

Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. J Stat Softw 33(1):1–22

Gaston A, Garcia-Vinas JI (2011) Modelling species distributions with penalised logistic regressions: a comparison with maximum entropy models. Ecol Model 222(13):2037–2041. https://doi.org/10.1016/j.ecolmodel.2011.04.015

George TS (2000) Varied thrush (Ixoreus naevius). In: Poole A (ed) The birds of North America Online. Cornell University, Ithaca, NY

Gill JA, Norris K, Potts PM, Gunnarsson TG, Atkinson PW, Sutherland WJ (2001) The buffer effect and large-scale population regulation in migratory birds. Nature 412(6845):436–438

Graham CH, Ferrier S, Huettman F, Moritz C, Peterson AT (2004) New developments in museum-based informatics and applications in biodiversity analysis. Trends Ecol Evol 19(9):497–503. https://doi.org/10.1016/j.tree.2004.07.006

Grinnell J (1917) The niche-relationships of the California Thrasher. Auk 34:427–433

Guillera-Arroita G (2017) Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. Ecography 40(2). https://doi.org/10.1111/ecog.02445

Guillera-Arroita G, Lahoz-Monfort JJ, Elith J (2014) Maxent is not a presence-absence method: a comment on Thibaud et al. Methods Ecol Evol 5(11):1192–1197. https://doi.org/10.1111/2041-210x.12252

Guillera-Arroita G, Lahoz-Monfort JJ, van Rooyen AR, Weeks AR, Tingley R (2017) Dealing with false-positive and false-negative errors about species occurrence at multiple levels. Methods Ecol Evol 8(9):1081–1091. https://doi.org/10.1111/2041-210x.12743

Guisan A, Harrell FE (2000) Ordinal response regression models in ecology. J Veg Sci 11 (5):617–626

Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. Ecol Lett 8(9):993–1009

Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. Ecol Model 135(2–3):147–186

Guisan A, Edwards TC, Hastie T (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol Model 157(2–3):89–100

Guisan A, Zimmermann NE, Elith J, Graham CH, Phillips S, Peterson AT (2007) What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? Ecol Monogr 77(4):615–630

Guisan A, Tingley R, Baumgartner JB, Naujokaitis-Lewis I, Sutcliffe PR, Tulloch AIT, Regan TJ, Brotons L, McDonald-Madden E, Mantyka-Pringle C, Martin TG, Rhodes JR, Maggini R, Setterfield SA, Elith J, Schwartz MW, Wintle BA, Broennimann O, Austin M, Ferrier S, Kearney MR, Possingham HP, Buckley YM (2013) Predicting species distributions for conservation decisions. Ecol Lett 16(12):1424–1435. https://doi.org/10.1111/ele.12189

Guisan A, Thuiller W, Zimmermann NE (2017) Habitat suitability and distribution models: applications with R. Cambridge University Press, Cambridge, UK

Hanski K, Ovaskainen O (2003) Metapopulation theory for fragmented landscapes. Theor Popul Biol 64(1):119–127. https://doi.org/10.1016/s0040-5809(03)00022-4

Hastie T, Fithian W (2013) Inference from presence-only data; the ongoing controversy. Ecography 36(8):864–867. https://doi.org/10.1111/j.1600-0587.2013.00321.x

Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference, and prediction, 2nd edn. Springer, New York

Hefley TJ, Broms KM, Brost BM, Buderman FE, Kay SL, Scharf HR, Tipton JR, Williams PJ, Hooten MB (2017) The basis function approach for modeling autocorrelation in ecological data. Ecology 98(3):632–646. https://doi.org/10.1002/ecy.1674

Hijmans RJ (2012) Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. Ecology 93(3):679–688

Hijmans RJ, Phillips S, Leathwick J, Elith J (2017) dismo: species distribution modeling. R package version 1.1.-4

Hirzel AH, Le Lay G (2008) Habitat suitability modelling and niche theory. J Appl Ecol 45 (5):1372–1381. https://doi.org/10.1111/j.1365-2664.2008.01524.x

Hirzel AH, Posse B, Oggier PA, Crettenand Y, Glenz C, Arlettaz R (2004) Ecological requirements of reintroduced species and the implications for release policy: the case of the bearded vulture. J Appl Ecol 41(6):1103–1116. https://doi.org/10.1111/j.0021-8901.2004.00980.x

Hirzel AH, Le Lay G, Helfer V, Randin C, Guisan A (2006) Evaluating the ability of habitat suitability models to predict species presences. Ecol Model 199(2):142–152. https://doi.org/10.1016/j.ecolmodel.2006.05.017

Holt RD (2009) Bringing the Hutchinsonian niche into the 21st century: ecological and evolutionary perspectives. Proc Natl Acad Sci U S A 106:19659–19665. https://doi.org/10.1073/pnas.0905137106

Hutchinson GE (1957) Concluding remarks. Population studies: animal ecology and demography. Cold Spring Harb Symp Quant Biol 22:415–427

Hutto RL, Young JS (2002) Regional landbird monitoring: perspectives from the Northern Rocky Mountains. Wildl Soc Bull 30(3):738–750

James FC, Johnston RF, Wamer NO, Niemi GJ, Boecklen WJ (1984) The grinnellian niche of the wood thrush. Am Nat 124(1):17–30. https://doi.org/10.1086/284250

Jimenez-Valverde A, Peterson AT, Soberon J, Overton JM, Aragon P, Lobo JM (2011) Use of niche models in invasive species risk assessments. Biol Invasions 13(12):2785–2797. https://doi.org/10.1007/s10530-011-9963-4

Johnson DS (2015) stocc: fit a spatial occupancy model via Gibbs sampling. R package version 1.30

Kadmon R, Farber O, Danin A (2004) Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. Ecol Appl 14(2):401–413

Kearney M, Porter W (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. Ecol Lett 12(4):334–350. https://doi.org/10.1111/j.1461-0248.2008.01277.x

Koshkina V, Wang Y, Gordon A, Dorazio RM, White M, Stone L (2017) Integrated species distribution models: combining presence-background data and site-occupany data with imperfect detection. Methods Ecol Evol 8(4):420–430. https://doi.org/10.1111/2041-210x.12738

Lahoz-Monfort JJ, Guillera-Arroita G, Wintle BA (2014) Imperfect detection impacts the performance of species distribution models. Glob Ecol Biogeogr 23(4):504–515. https://doi.org/10.1111/geb.12138

Lawson CR, Hodgson JA, Wilson RJ, Richards SA (2014) Prevalence, thresholds and the performance of presence-absence models. Methods Ecol Evol 5(1):54–64. https://doi.org/10.1111/2041-210x.12123

Liaw A, Wiener M (2002) Classification and regression by randomforest. R News 2(3):18–22

Lira-Noriega A, Soberon J, Miller CP (2013) Process-based and correlative modeling of desert mistletoe distribution: a multiscalar approach. Ecosphere 4(8):99. https://doi.org/10.1890/es13-00155.1

Liu CR, Berry PM, Dawson TP, Pearson RG (2005) Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28(3):385–393

Liu C, White M, Newell G (2013) Selecting thresholds for the prediction of species occurrence with presence-only data. J Biogeogr 40(4):778–789. https://doi.org/10.1111/jbi.12058

Lobo JM, Jimenez-Valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models. Glob Ecol Biogeogr 17(2):145–151. https://doi.org/10.1111/j.1466-8238.2007.00358.x

Loiselle BA, Jorgensen PM, Consiglio T, Jimenez I, Blake JG, Lohmann LG, Montiel OM (2008) Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? J Biogeogr 35(1):105–116. https://doi.org/10.1111/j.1365-2699.2007.01779.x

Lutolf M, Kienast F, Guisan A (2006) The ghost of past species occurrence: improving species distribution models for presence-only data. J Appl Ecol 43(4):802–815. https://doi.org/10.1111/j.1365-2664.2006.01191.x

MacKenzie DI, Nichols JD, Lachman GB, Droege S, Royle JA, Langtimm CA (2002) Estimating site occupancy rates when detection probabilities are less than one. Ecology 83(8):2248–2255

MacKenzie DI, Nichols JD, Hines JE, Knutson MG, Franklin AB (2003) Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. Ecology 84 (8):2200–2207

MacKenzie DI, Nichols JD, Royle JA, Pollock KH, Bailey LL, Hines JE (2006) Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence. Elsevier, Amsterdam

Manly BFJ, McDonald LL, Thomas DL, McDonald TL, Erickson WP (2002) Resource selection by animals: statistical design and analysis for field studies. Kluwer Academic Publishers, Dordrecht, the Netherlands

Marmion M, Parviainen M, Luoto M, Heikkinen RK, Thuiller W (2009) Evaluation of consensus methods in predictive species distribution modelling. Divers Distrib 15(1):59–69. https://doi.org/10.1111/j.1472-4642.2008.00491.x

Martin Y, Van Dyck H, Dendoncker N, Titeux N (2013) Testing instead of assuming the importance of land use change scenarios to model species distributions under climate change. Glob Ecol Biogeogr 22(11):1204–1216. https://doi.org/10.1111/geb.12087

Martinez-Meyer E, Peterson AT, Servin JI, Kiff LF (2006) Ecological niche modelling and prioritizing areas for species reintroductions. Oryx 40(4):411–418. https://doi.org/10.1017/s0030605306001360

McCarthy KP, Fletcher RJ, Rota CT, Hutto RL (2012) Predicting species distributions from samples collected along roadsides. Conserv Biol 26(1):68–77. https://doi.org/10.1111/j.1523-1739.2011.01754.x

Merow C, Smith MJ, Silander JA (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography 36(10):1058–1069. https://doi.org/10.1111/j.1600-0587.2013.07872.x

Miller JA, Holloway P (2015) Incorporating movement in species distribution models. Prog Phys Geogr 39(6):837–849. https://doi.org/10.1177/0309133315580890

Miller DA, Nichols JD, McClintock BT, Grant EHC, Bailey LL, Weir LA (2011) Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. Ecology 92(7):1422–1428. https://doi.org/10.1890/10-1396.1

Naimi B, Araújo MB (2016) sdm: a reproducible and extensible R platform for species distribution modelling. Ecography 39(4):368–375. https://doi.org/10.1111/ecog.01881

Norris K (2004) Managing threatened species: the ecological toolbox, evolutionary theory and declining-population paradigm. J Appl Ecol 41(3):413–426

Olden JD, Lawler JJ, Poff NL (2008) Machine learning methods without tears: a primer for ecologists. Q Rev Biol 83(2):171–193. https://doi.org/10.1086/587826

Ovaskainen O, Hottola J, Siitonen J (2010) Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. Ecology 91(9):2514–2521. https://doi.org/10.1890/10-0173.1

Pacifici K, Reich BJ, Miller DAW, Gardner B, Stauffer G, Singh S, McKerrow A, Collazo JA (2017) Integrating multiple data sources in species distribution modeling: a framework for data fusion. Ecology 98(3):840–850. https://doi.org/10.1002/ecy.1710

Pearce J, Ferrier S (2000) Evaluating the predictive performance of habitat models developed using logistic regression. Ecol Model 133(3):225–245

Pearson RG, Dawson TP (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Glob Ecol Biogeogr 12(5):361–371. https://doi.org/10.1046/j.1466-822X.2003.00042.x

Peterson AT (2003) Predicting the geography of species' invasions via ecological niche modeling. Q Rev Biol 78(4):419–433. https://doi.org/10.1086/378926

Peterson AT, Soberon J (2012) Species distribution modeling and ecological niche modeling: getting the concepts right. Natureza & Conservacao 10(2):102–107. https://doi.org/10.4322/natcon.2012.019

Peterson AT, Papes M, Soberon J (2008) Rethinking receiver operating characteristic analysis applications in ecological niche modeling. Ecol Model 213(1):63–72. https://doi.org/10.1016/j.ecolmodel.2007.11.008

Peterson AT, Soberon J, Pearson RG, Anderson RP, Martinez-Mery E, Nakamura M, Araújo MB (2011) Ecological niches and geographic distributions. Princeton University Press, Princeton, NJ

Phillips SJ, Dudik M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31(2):161–175. https://doi.org/10.1111/j.0906-7590.2008.5203.x

Phillips SJ, Elith J (2010) POC plots: calibrating species distribution models with presence-only data. Ecology 91(8):2476–2484. https://doi.org/10.1890/09-0760.1

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Model 190(3–4):231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Phillips SJ, Dudik M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol Appl 19(1):181–197

Phillips SJ, Anderson RP, Dudik M, Schapire RE, Blair ME (2017) Opening the black box: an open-source release of Maxent. Ecography 40(7):887–893. https://doi.org/10.1111/ecog.03049

Plath M, Moser C, Bailis R, Brandt P, Hirsch H, Klein AM, Walmsley D, von Wehrden H (2016) A novel bioenergy feedstock in Latin America? Cultivation potential of Acrocomia aculeata under current and future climate conditions. Biomass Bioenergy 91:186–195. https://doi.org/10.1016/j.biombioe.2016.04.009

Potts JM, Elith J (2006) Comparing species abundance models. Ecol Model 199(2):153–163. https://doi.org/10.1016/j.ecolmodel.2006.05.025

Prasad AM, Iverson LR, Liaw A (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9(2):181–199. https://doi.org/10.1007/s10021-005-0054-1

Pulliam HR (1988) Sources, sinks, and population regulation. Am Nat 132(5):652–661

Pulliam HR (2000) On the relationship between niche and distribution. Ecol Lett 3(4):349–361

Raxworthy CJ, Martinez-Meyer E, Horning N, Nussbaum RA, Schneider GE, Ortega-Huerta MA, Peterson AT (2003) Predicting distributions of known and unknown reptile species in Madagascar. Nature 426(6968):837–841. https://doi.org/10.1038/nature02205

Real LA, Brown JH (eds) (1991) Foundations of ecology: classic papers with commentaries. University of Chicago Press, Chicago

Renner IW, Warton DI (2013) Equivalence of MAXENT and poisson point process models for species distribution modeling in ecology. Biometrics 69(1):274–281. https://doi.org/10.1111/j.1541-0420.2012.01824.x

Renner IW, Elith J, Baddeley A, Fithian W, Hastie T, Phillips SJ, Popovic G, Warton DI (2015) Point process models for presence-only analysis. Methods Ecol Evol 6(4):366–379. https://doi.org/10.1111/2041-210x.12352

Robertson BA, Hutto RL (2006) A framework for understanding ecological traps and an evaluation of existing evidence. Ecology 87(5):1075–1085

Rodrigues ASL, Akcakaya HR, Andelman SJ, Bakarr MI, Boitani L, Brooks TM, Chanson JS, Fishpool LDC, Da Fonseca GAB, Gaston KJ, Hoffmann M, Marquet PA, Pilgrim JD, Pressey RL, Schipper J, Sechrest W, Stuart SN, Underhill LG, Waller RW, Watts MEJ, Yan X (2004) Global gap analysis: priority regions for expanding the global protected-area network. Bioscience 54(12):1092–1100. https://doi.org/10.1641/0006-3568(2004)054[1092:ggaprf]2.0.co;2

Rota CT, Fletcher RJ Jr, Evans JM, Hutto RL (2011) Does accounting for detectability improve species distribution models. Ecography 34:659–670

Rotenberry JT, Wiens JA (1980) Habitat structure, patchiness, and avian communities in North-American steppe vegetation: a multivariate-analysis. Ecology 61(5):1228–1250. https://doi.org/10.2307/1936840

Rotenberry JT, Preston KL, Knick ST (2006) Gis-based niche modeling for mapping species' habitat. Ecology 87(6):1458–1464

Royle JA, Chandler RB, Yackulic C, Nichols JD (2012) Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. Methods Ecol Evol 3 (3):545–554. https://doi.org/10.1111/j.2041-210X.2011.00182.x

Schlaepfer MA, Runge MC, Sherman PW (2002) Ecological and evolutionary traps. Trends Ecol Evol 17(10):474–480

Scott JM, Davis F, Csuti B, Noss R, Butterfield B, Groves C, Anderson H, Caicco S, Derchia F, Edwards TC, Ulliman J, Wright RG (1993) GAP analysis: a geographic approach to protection of biological diversity. Wildl Monogr (123):1–41

Smolik MG, Dullinger S, Essl F, Kleinbauer I, Leitner M, Peterseil J, Stadler LM, Vogl G (2010) Integrating species distribution models and interacting particle systems to predict the spread of an invasive alien plant. J Biogeogr 37(3):411–422. https://doi.org/10.1111/j.1365-2699.2009.02227.x

Soberón J (2007) Grinnellian and Eltonian niches and geographic distributions of species. Ecol Lett 10(12):1115–1123. https://doi.org/10.1111/j.1461-0248.2007.01107.x

Soberón JM (2010) Niche and area of distribution modeling: a population ecology perspective. Ecography 33(1):159–167. https://doi.org/10.1111/j.1600-0587.2009.06074.x

Soberón J, Nakamura M (2009) Niches and distributional areas: concepts, methods, and assumptions. Proc Natl Acad Sci U S A 106:19644–19650. https://doi.org/10.1073/pnas.0901637106

Soberón J, Peterson AT (2005) Interpretation of models of fundamental ecological niches and species' distributional areas. Biodivers Inform 2:1–10

Tewksbury JJ, Garner L, Garner S, Lloyd JD, Saab V, Martin TE (2006) Tests of landscape influence: nest predation and brood parasitism in fragmented ecosystems. Ecology 87 (3):759–768

Thomas CD, Cameron A, Green RE, Bakkenes M, Beaumont LJ, Collingham YC, Erasmus BFN, de Siqueira MF, Grainger A, Hannah L, Hughes L, Huntley B, van Jaarsveld AS, Midgley GF, Miles L, Ortega-Huerta MA, Peterson AT, Phillips OL, Williams SE (2004) Extinction risk from climate change. Nature 427(6970):145–148. https://doi.org/10.1038/nature02121

Thuiller W, Georges D, Engler R, Breiner F (2016) biomod2: ensemble platform for species distribution modeling. R package version 3.3.-7

Tibshirani R (1996) Regression shrinkage and selection via the Lasso. J R Stat Soc Series B Methodol 58(1):267–288

Urbanek S (2017) rJava: low-level R to Java interface. R package 0.9-9

Vallecillo S, Brotons L, Thuiller W (2009) Dangers of predicting bird species distributions in response to land-cover changes. Ecol Appl 19(2):538–549. https://doi.org/10.1890/08-0348.1

Van Horne B (1983) Density as a misleading indicator of habitat quality. J Wildl Manag 47:893–901

VanDerWal J, Shoo LP, Graham C, William SE (2009) Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? Ecol Model 220 (4):589–594. https://doi.org/10.1016/j.ecolmodel.2008.11.010

Vielledent G, Merow C, Guelat J, Latimer AM, Kery M, Gelfand AE, Wilson AM, F. Mortier, Silander Jr JA (2014) hSDM: hierachical Bayesian species distribution models. R package version 1.4

Ward G, Hastie T, Barry S, Elith J, Leathwick JR (2009) Presence-only data and the EM algorithm. Biometrics 65(2):554–563. https://doi.org/10.1111/j.1541-0420.2008.01116.x

Warton DI, Shepherd LC (2010) Poisson point process models solve the “pseudo-absence problem” for presence-only data in ecology. Ann Appl Stat 4(3):1383–1402. https://doi.org/10.1214/10-aoas331

Wenger SJ, Olden JD (2012) Assessing transferability of ecological models: an underappreciated aspect of statistical validation. Methods Ecol Evol 3(2):260–267. https://doi.org/10.1111/j.2041-210X.2011.00170.x

Whittaker RH, Levin SA, Root RB (1973) Niche, habitat, and ecotope. Am Nat 107(955):321–338. https://doi.org/10.1086/282837

Wickham H (2009) ggplot2: elegant graphics for data analysis. Springer-Verlag, New York

Wiens JJ, Ackerly DD, Allen AP, Anacker BL, Buckley LB, Cornell HV, Damschen EI, Davies TJ, Grytnes JA, Harrison SP, Hawkins BA, Holt RD, McCain CM, Stephens PR (2010) Niche conservatism as an emerging principle in ecology and conservation biology. Ecol Lett 13 (10):1310–1324. https://doi.org/10.1111/j.1461-0248.2010.01515.x

Williams JW, Jackson ST (2007) Novel climates, no-analog communities, and ecological surprises. Front Ecol Environ 5(9):475–482. https://doi.org/10.1890/070037

Wilson KA, Westphal MI, Possingham HP, Elith J (2005) Sensitivity of conservation planning to different approaches to using predicted species distribution data. Biol Conserv 122(1):99–112. https://doi.org/10.1016/j.biocon.2004.07.004

Wisz MS, Pottier J, Kissling WD, Pellissier L, Lenoir J, Damgaard CF, Dormann CF, Forchhammer MC, Grytnes JA, Guisan A, Heikkinen RK, Hoye TT, Kuhn I, Luoto M, Maiorano L, Nilsson MC, Normand S, Ockinger E, Schmidt NM, Termansen M, Timmermann A, Wardle DA, Aastrup P, Svenning JC (2013) The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. Biol Rev 88 (1):15–30. https://doi.org/10.1111/j.1469-185X.2012.00235.x

Wood SN (2006) Generalized additive models: an introduction with R. Chapman and Hall and CRC, Boca Raton, FL

Yackulic CB, Chandler R, Zipkin EF, Royle JA, Nichols JD, Grant EHC, Veran S (2013) Presence-only modelling using MAXENT: when can we trust the inferences? Methods Ecol Evol 4 (3):236–243. https://doi.org/10.1111/2041-210x.12004

Yackulic CB, Nichols JD, Reid J, Der R (2015) To predict the niche, model colonization and extinction. Ecology 96(1):16–23. https://doi.org/10.1890/14-1361.1