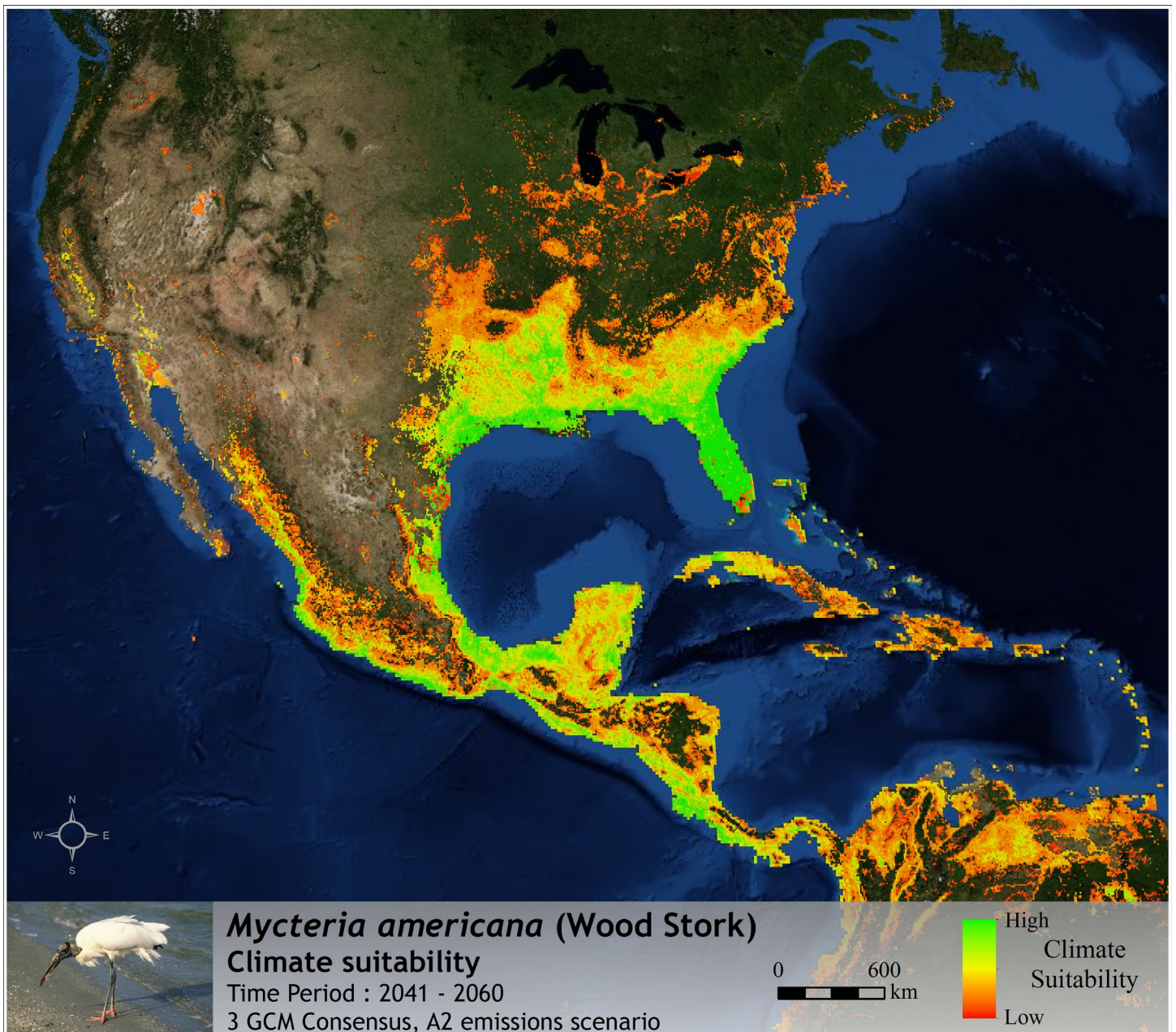# Use and Interpretation of Climate Envelope Models: A Practical Guide

James I. Watling, University of Florida

Laura A. Brandt, U.S. Fish and Wildlife Service

Frank J. Mazzotti, University of Florida

Stephanie S. Romañach, U.S. Geological Survey

Mycteria americana (Wood Stork)
Climate suitability
Time Period : 2041 - 2060
3 GCM Consensus, A2 emissions scenario

0   600
km

High
Climate
Suitability
Low

Preface

This guidebook is intended to provide a practical overview of climate envelope modeling for conservation professionals and natural resource managers. The material is intended for people with little background or experience in climate envelope modeling who want to better understand and interpret models developed by others and the results generated by such models, or want to do some modeling themselves. This is not an exhaustive review of climate envelope modeling, but rather a brief introduction to some key concepts in the discipline. Readers interested in a more in-depth treatment of much of the material presented here are referred to an excellent book, *Mapping Species Distributions: Spatial Inference and Prediction* by Janet Franklin. Also, a recent review (Araújo & Peterson 2012) provides an excellent, though more technical, discussion of many of the issues dealt with here. Here we treat selected topics from a practical perspective, using minimal jargon to explain and illustrate some of the many issues that one has to be aware of when using climate envelope models. When we do introduce specialized terminology in the guidebook, we bold the term when it is first used; a glossary of these terms is included at the back of the guidebook.

We have written the guidebook from the perspective of a natural resource manager attending a national conference, reading scientific articles or viewing webinars on climate change science related to species and habitat. That manager would likely read or hear about research from across the United States or the world, including many talks in which different groups are using climate envelope models to forecast species responses to future climate change. As is often the case in real life, the different research groups presenting at the conference may be addressing broadly similar questions, but the details of their approaches differ. Some may be working with endangered species, others with invasive species, and some with migratory species. Some researchers may work with plants, and others with animals (one of the great things about climate envelope modeling is that the approach is the same regardless of the species of interest). They probably use different methods to create their models, the spatial extent of their modeling may vary, and they may use different approaches to validate their models. How would a non-expert compare such seemingly disparate models? Are there 'best practices' to keep in mind when evaluating climate envelope models? How would you know if someone is making a reasonable inference, or violating a major assumption of the modeling? We attempt to deal with some of these issues and other important concepts related to climate envelope modeling throughout the guidebook.

The guidebook is structured as a series of questions that might result from hearing about work with climate envelope models at a conference, webinar, or in the primary literature. We begin with some general aspects of models and relevant background information. We then examine different components of the models, including species and climate data. We illustrate key concepts with examples from our own work developing climate envelope models for Florida's threatened and endangered species, as well as work of others in the field of distribution modeling.

We urge readers of this guidebook to keep in mind that models are one part of a natural resource manager's toolbox, but they are not the only, or even the most important tool. It can't be repeated enough that 'all models are wrong, but some models are useful'. Models are abstractions of reality and cannot, by definition, include all the relevant or interesting parts of a system. Models can help guide our thinking by describing plausible outcomes of different conditions or management decisions and they help to suggest hypotheses to be tested with additional research. Models do not provide all the answers and should not be accepted uncritically—all models make simplifying assumptions and abstract reality in different ways. It is our hope that this guidebook helps identify some of the key assumptions and uncertainties underlying climate envelope modeling while providing useful insights that will help users interpret the results of climate envelope models.

Contact information for authors:
- James I. Watling, Ft Lauderdale Research and Education Center, 3205 College Avenue, Ft Lauderdale, FL 33314 USA. E-mail: watlingj@ufl.edu
- Laura A. Brandt, U. S. Fish and Wildlife Service, Ft Lauderdale Research and Education Center, 3205 College Avenue, Ft Lauderdale, FL 33314 USA. E-mail: laura_brandt@fws.gov
- Frank J. Mazzotti, Ft Lauderdale Research and Education Center, 3205 College Avenue, Ft Lauderdale, FL 33314 USA. E-mail: fjma@ufl.edu
- Stephanie S. Romañach, U. S. Geological Survey, Southeast Ecological Science Center, 3205 College Avenue, Ft Lauderdale, FL 33314 USA. E-mail: sromanach@usgs.gov

Suggested citation: Watling JI, LA Brandt, FJ Mazzotti & SS Romañach. 2013. Use and Interpretation of Climate Envelope Models: A Practical Guide. University of Florida, 43 pp.

For comments or questions on any of the material in this document, please contact James Watling at the University of Florida, watlingj@ufl.edu.

Table of Contents

What is a climate envelope model?

Throughout this document we use the term **'climate envelope model'** to refer to a subset of **species distribution models** that use climate variables to make spatial predictions of environmental suitability for a species. Species distribution models (also called 'niche models') use rules or mathematical functions to describe associations between species occurrence and environmental conditions. On the basis of these rules, users can extrapolate an index of environmental suitability or probability of occurrence for the modeled species from a map of environmental conditions. This **extrapolation** involves taking the mathematical functions describing species-environment associations and applying those same functions to environmental data from some other time or place. The variables used in species distribution models may include climate, land cover, topography or any other variable relevant to the species being modeled. Species distribution models may be used for a variety of different purposes, such as identifying species diversity hotspots (Platts et al. 2010) and predicting potential ranges of invasive species (Bidinger et al. 2012), or forecasting climate change effects on biodiversity (LaSorte & Jetz 2012). In a strict sense, climate envelope models refer to a particular type of model that defines minimum and maximum values of climate boundaries around species occurrences, thereby delimiting a 'climate envelope' within which species occur. In this guide, however, we use the term more generally to refer to any type of species distribution model that draws on climate data to define environmental suitability.

What are climate envelope models designed to do?

Climate envelope models delineate areas of climate suitability for plant or animal species of interest by correlating georeferenced species occurrences (presences and absences) with observed climate conditions at occurrence sites. Climate envelope models are spatially-explicit, and modelers generally use maps of species occurrences and climate conditions to describe species-climate relationships within individual grid cells of those maps. According to our definition, climate envelope models only include climate variables, so the models are only describing areas (grid cells) where climate is suitable (or projected to become suitable) for the species being modeled. There are many other factors in addition to climate that may limit species distributions, such as habitat availability, habitat fragmentation, competition with other species, and predators. Some of these factors can be included along with climate variables using the same modeling framework (variables such as land cover and elevation), but that would constitute a more general species distribution model. Here we focus specifically on climate variables, and climate envelope models. Climate envelope models do not explicitly incorporate species traits or additional information into models. Some species information is introduced into models *implicitly*, during the variable selection process.

We address the theoretical and practical implications of variable selection below. Later in the document we also discuss in more detail what a climate envelope model does not do, and make

suggestions for judicious interpretation of climate envelope model outputs. Remember, though, that models are abstractions of reality, and should not be used as the sole basis for natural resource decision-making, but rather serve as one of many tools that can help inform decisions.

Climate envelope models generally define climate suitability first based on contemporary climate, although the exact time period used to define 'contemporary' conditions varies among studies. Also note that climate envelope models are not necessarily defined first for the contemporary period and then extrapolated to the future. Although this is a very common way of doing climate envelope modeling, one can just as easily define a model for some past time period, and extrapolate the model to the modern day, or define a model using modern climate and extrapolate it to the past. However, for the purposes of understanding potential climate change effects, scientists generally calibrate models based on twentieth century climate, or some subset (many studies use climate 'normals', or long-term averages, from a 30—50 year period at the end of the twentieth century).

Throughout the guidebook, we refer to **calibration** as the process of fitting the mathematical species-climate association, which is followed by the use of evaluation metrics to describe how well the model links species occurrences to climate (this process is described in more detail later). Most of the time, scientists are interested in extrapolating the contemporary species-climate relationship into the future to describe possible effects of climate change on species. In that case, the mathematical relationship between species occurrences and climate described for the present day is assumed to hold true in the future, and areas where climate is expected to be suitable for

a species in the future are identified using descriptions of future climate conditions (Figure 1).

What are the limitations of climate envelope models?

The first thing to remember is that a climate envelope model is only using information on climate to establish a mathematical association between species occurrences and climate variables. Climate envelope models make broad generalizations about areas that may have suitable climate for species, but that's it. Climate does not generally vary on the scale of 50—100 m, or even a few kilometers, so climate envelope models are not likely to be useful for making predictions about where species will be located within a typical wildlife refuge, for example. Climate suitability alone does not guarantee the species will actually be present in an area.

Because climate envelope models do not incorporate species-specific data other than occurrence, models do not predict traits such as changes in phenology (the timing of organism's life cycles, such as egg laying or flowering times) or demographic rates (births, deaths, immigration or emigration). The modeling of such responses falls under the umbrella of **mechanistic models**, physiological models, or process-based models (Kearney & Porter 2009). Studies that have compared predictions from climate envelope models to predictions from mechanistic models that incorporate demographic and physiological data suggest that outputs from these different types of models are not as discrepant as may be expected given the differences in the input data. One study (Kearney et al. 2010) found that results were broadly comparable. Another study (Buckley et al. 2010) found that predictions from both climate envelope and mechanistic models alone were not as successful in predicting occurrences
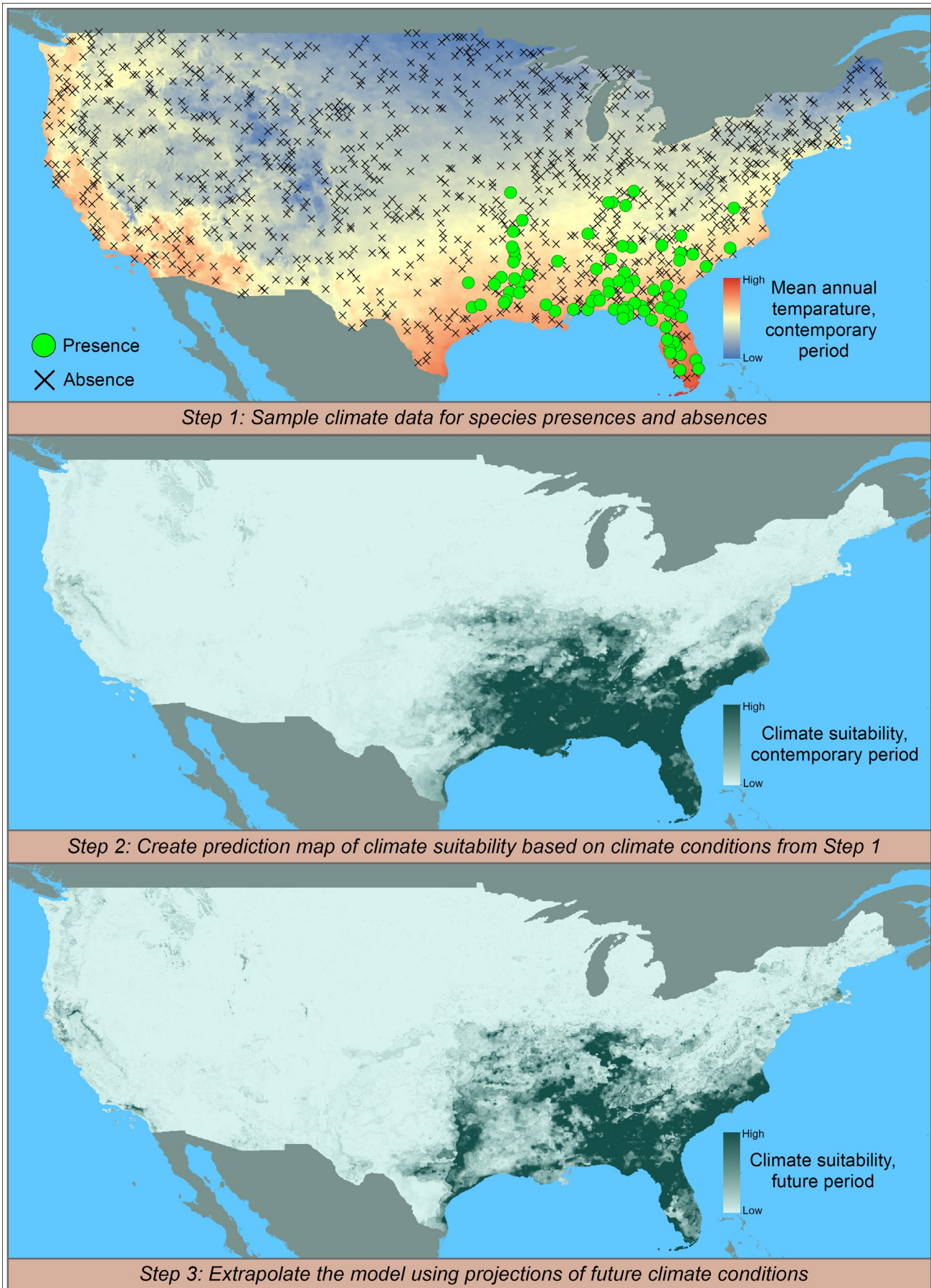
Figure 1. Conceptual overview of the climate envelope modeling process. In step one, climate data are compiled at sites where a species exists, as well as sites where species are absent or their status is unknown. In step two, a mathematical equation is used to extrapolate an estimate of climate suitability for the species using the same climate data as in step one. In step three, using the same species-climate relationship as in step two, climate suitability is extrapolated to a new (future) period based on future climate conditions.

of the Eastern fence lizard (*Sceloporous undulatus*) as a hybrid approach combining elements of both modeling approaches. Recently-developed extensions of mechanistic modeling such as dynamic range models (Pagel & Schurr 2012; Shurr et al. 2012) focus special attention on dispersal, an important characteristic that is missing entirely from traditional climate envelope models. Because data inputs for climate envelope models are much less demanding than for mechanistic models, we expect that climate envelope modeling will continue to be used for modeling of species for which detailed demographic or physiological data are lacking, even as approaches for mechanistic modeling become more user-friendly. We reiterate, however, that models of any type should only be part of the decision-making toolbox, and not the sole basis for conservation planning.

What are the differences between climate envelope and mechanistic models, and what are you likely to learn from each of them? In general, mechanistic models are more data intensive, because they require detailed information on how individual fitness varies as a function of climate (e.g., survival growth and reproductive output under different temperature or precipitation conditions). These data are generally used to fit specific equations describing things like physiological response curves. Since those types of data are not available for many species, physiological models are available for only a few species. In contrast, climate envelope models can be created for many different species (e.g., almost 3000 species of terrestrial vertebrates in the North and South America, Lawler et al. 2009). Because the outputs of mechanistic models include information on how fitness traits vary across climate gradients, they can provide insights into the effects of climate on specific aspects of an organism's life history in a presentation or paper on mechanistic models than in a presentation on climate envelope models.

However, because climate envelope models can be easily modified to include data on other types of environmental conditions (land cover, elevation, etc), you are more likely to learn about the relative importance of climate in determining species range limits or range shifts when reading a paper on climate envelope models than in a paper on mechanistic models. In general, it probably makes sense to think about climate envelope models as being focused on broad patterns and limiting factors for species, often in the context of a relatively coarse-filtered screening tool for evaluating species susceptibility to climate change. Mechanistic models, on the other hand, give much more insight into how climate change affects demography of individual species and are probably more likely to be used for species already known to be at risk for negative effects of climate change or that are of particular economic or cultural importance (e.g., waterfowl and sport fish).

How are climate envelope models created?

Creating a climate envelope model is fairly straight forward in concept. A researcher gathers occurrence data and contemporary climate data to establish the relationships between species occurrence and climate variables. The relationship is calculated using one or more **algorithms**, and the researcher evaluates the results. Once the relationship has been determined, a researcher can use that relationship with projections of future climate to describe the future climate envelope for the species. The modeling process can vary quite a bit, depending on factors such as the algorithm and statistical software used for modeling and the extent to which input data are pre- or post-processed. Since it is not practical to include a detailed how-to guide that is applicable to all types of models and available platforms, here we highlight some of the key steps in the modeling process.

Model evaluation or validation is a critical part of model development, and has important implications for model interpretation. Scientists often use climate envelope modeling to make a projection of future conditions ('**prediction**' and '**projection**' have very specific and sometimes contentious definitions in the literature on climate change, but in general scientists use the word projection to talk about a conditional, far-off in the future condition, and prediction to refer to a more immediate, relatively verifiable condition; Bray & von Storch 2009). Since we cannot wait to see if species responses to climate change corroborate our model projections, how do we know if we have a 'good' model? In general, we evaluate the model based on how well it is able to classify *contemporary* presences and absences/**'pseudo-absences'**. It can be quite hard to determine true 'biological' absences due to challenges in surveying wild populations, particularly animals. Therefore, many users include 'pseudo-absences' in climate envelope models. This is discussed in more detail later. In other words, because we use the same functional relationship between species occurrence and climate to create a projection of future climate suitability as we do to create the contemporary model, we can evaluate model performance based on its ability to accurately say something about where species are or not based on today's climate. This is why we first create an extrapolation of climate suitability for the species being modeled based on contemporary climate—we want to create a map of climate suitability that we can evaluate using known occurrence data. Of course, the best way to evaluate a model would be to use two completely independent presence/absence datasets, the first one to calibrate the model, and another, completely independent one to evaluate the model. In reality, we often do not have independent presence/absence survey data for species, so we usually use a data partitioning or **cross-validation** procedure for model evaluation. There are a couple of different

ways to do that, both of which require dividing the occurrence data (all the data, both presences and absences) into '**training**' and '**testing**' subsets. The training data are used to calibrate a model, which is evaluated using the test subset. It is invariably true that the use of truly independent survey data for testing will result in lower model performance than when a cross-validation procedure is used. This lower estimate, however, is a better indicator of model performance than an inflated estimate from cross-validation. Because it is relatively rare that researchers have access to independent presence/absence data for model evaluation, we focus on cross-validation techniques below.

Researchers use a variety of cross-validation procedures to evaluate performance based on a model's ability to correctly differentiate presences and absences. One such technique is called *k*-**fold partitioning**, wherein the presence-absence dataset is divided into *k* number of groups (e.g., with 10 groups you have ten-fold partitioning, with five groups you have five-fold partitioning). All but one of the groups are combined and used to train the model, and the last group is used to test the model. The process is repeated *k* times, until each subset has been the test subset once, and part of the training subset *k*-1 times (Figure 2). **Subsetting** is a different type of cross-validation. In that case, a researcher subsets the species occurrence database (again including both presences and absences) into training and testing subsets. Usually, a modeler assigns about 75% of the occurrences to the training dataset and the remaining 25% are retained for validation, though slightly different proportions are sometimes used. The subsetting procedure is repeated a number of times (say 100) so that different data points comprise the training and testing partitions each time, and the average results are presented. Using both cross-validation procedures, the success of the model in

correctly characterizing presences and absences in the test dataset is examined (details of the metrics used to evaluate model performance are discussed in the next section). If model performance is acceptable, all the presence/absence data may be used in a final, 'real' model that can then be applied to contemporary climate conditions or used to make a projection of climate suitability under future conditions.

Does the type of cross-validation procedure matter for differentiating a good versus bad model? In short, probably not that much. Although we are unaware of studies that specifically compare cross validation techniques in climate envelope models,

both of the approaches described here are widely used in the literature. What is probably more important is repeating the subsetting process a number of times with random partitions of the data to reduce the chance that, by luck, a model tests particularly well, even if examination using another test subset (or fold) would result in much lower performance. Model evaluation is considered an integral part of the modeling process, and users should treat results with particular caution if the results of a partitioning or subsetting procedure are not explicitly presented along with the model's projections. Furthermore, the gold standard for model evaluation is the use of independent presence/absence data, and users

Examples of two repetitions of 10-fold partitioning (a,b) and sub-setting (c,d) cross-validation techniques. In (a,b), presences and absences illustrated have been randomly assigned to ten different groups indicated by numbers 1—10. In (a), the group marked "1" is used to validate a model built from all other groups combined, and in (b) group "2" is used. In (c,d), the random subsets are generated before each cross-validation, sub-setting 75% of the data for training ("T"), and 25% for validation ("V").

● Presence used in validation test
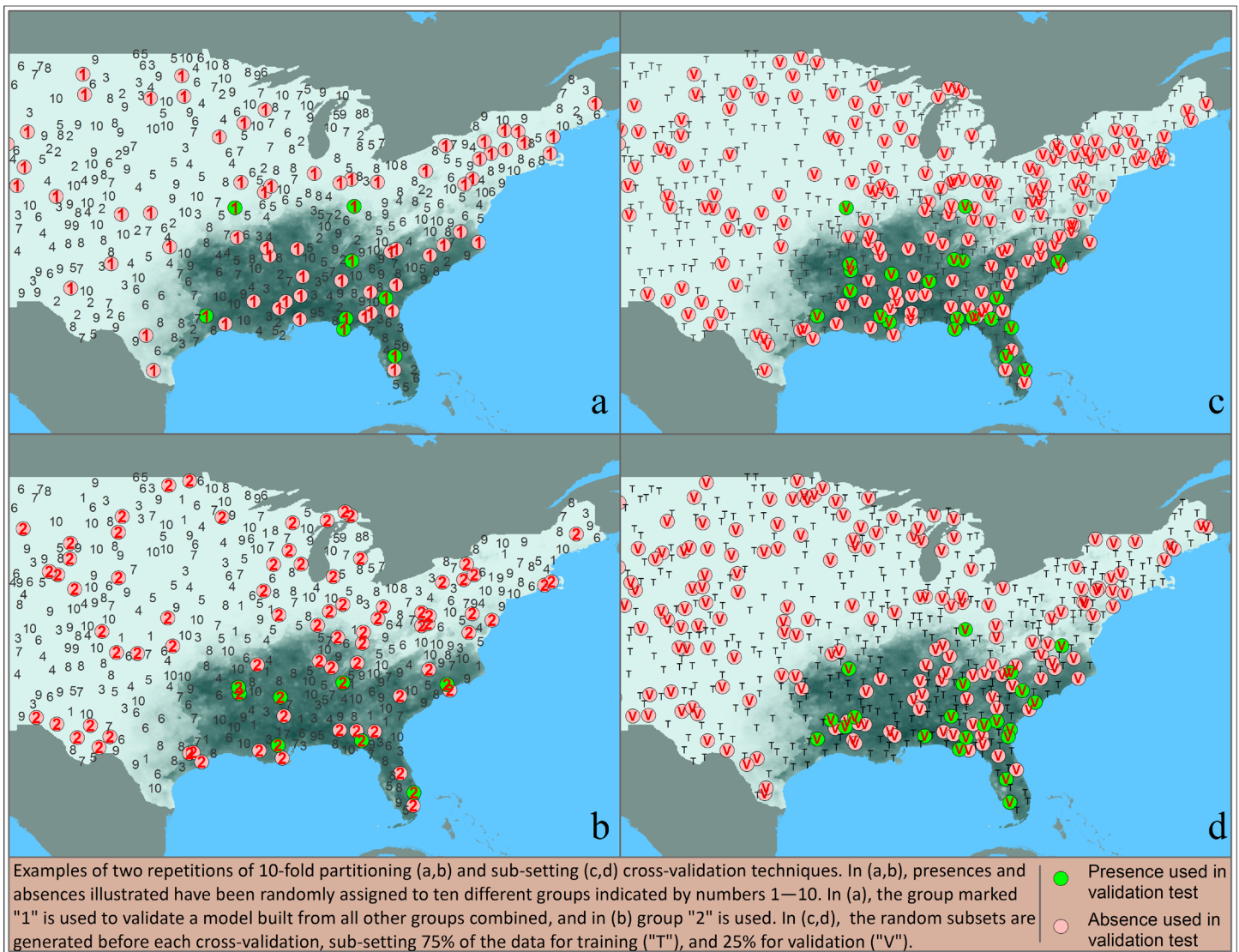○ Absence used in validation test

Figure 2. Example of 10-fold cross validation.

should almost always prefer information from models evaluated against independent survey data than from models evaluated using a cross-validation procedure.

How is model performance evaluated?

There are many ways to evaluate performance in climate envelope models. However, before going into some details about different ways to evaluate model performance, it's worth considering what we are evaluating. When people talk about model performance, they are talking about the model's ability to correctly distinguish sites where a species is known to occur from those where it is absent (in the case of true field survey-based absences) or its occupancy status is unknown (when using pseudo-absences). When using a climate envelope model to forecast future climate change effects, the only evaluation data we have is for species occurrences we have already observed (those from the contemporary period), so we only evaluate performance based on the model's ability to correctly distinguish occupied and unoccupied sites based on contemporary data. When it comes to forecasting future conditions, we really don't know if a model is good or not, we just assume that if a model did a good job classifying contemporary occurrences, it will do a good job in the future.

There are two basic types of errors to consider when evaluating model performance: omission and commission error (Figure 3). **Omission error** occurs when a model fails to predict presence (or climate suitability) in an area where the species really does occur. In other words, the model omits a known species presence. **Commission error** happens when the model predicts presence/suitability in an area where the species is absent or occupancy is unknown. Commission error is usually interpreted as an **overprediction**, although that may not be entirely

accurate because of the uncertain nature of absences (Lobo et al. 2010).

In general, modelers distinguish between two different types of evaluation metrics: threshold-dependent and threshold-independent metrics. The outputs of a climate envelope model are generally interpreted as estimates of climate suitability or probabilities of species occurrence ranging from 0—1. **Threshold-independent metrics** evaluate model performance using only the probabilities resulting from the model. In other words, threshold-independent metrics evaluate model performance using the 'raw' probabilities of climate suitability for each grid cell in a prediction map: 0.01, 0.23, 0.79 and so on, with higher numbers indicating greater suitability. **Threshold-dependent metrics** require the user to convert raw probabilities into two categories by identifying a threshold above which probabilities are interpreted as representing areas of 'suitable' climate (and generally coded with a '1'), and below which probabilities are interpreted as representing 'unsuitable' areas (generally coded with a '0', Table 1). There is a well-developed literature on the selection of appropriate thresholds; the topic is a detailed one so we refer readers to Freeman & Moisen (2008) for a detailed discussion of threshold selection in species distribution models.

Table 1. Examples of how different probabilities may be converted to a categorical (0/1) prediction differentiating suitable (coded as 1) and unsuitable (coded as 0) areas. If the threshold is 0.50 and a cell probability is 0.23, that cell would be classified as 'unsuitable' and coded as a 0, whereas a cell with probability 0.56 would be classified as 'suitable'. If the threshold is raised to 0.75, both cells would be classified as 'unsuitable'.

	Probability			
	0.23	0.50	0.56	0.75
Threshold of 0.50	0	1	1	1
Threshold of 0.75	0	0	0	1

There are many individual threshold-independent and threshold-dependent metrics available for model

Example of areas identified as "overprediction" (commission error; yellow "X"s) and "underprediction" (ommission error; red circles) by model performance metrics. Here, the threshold was applied at a value that roughly balanced specificity and sensitivity.

○ Presence
✕ Absence

Prob. Threshold = 0.45
Sensitivity = 0.86
(70 / 81)
Specificity = 0.88
(880 / 1000)
Kappa = 0.46

Here, the threshold was applied so as to maximize sensitivity. All occurences fall within areas predicted as suitable; however, this comes at the cost of model specificity (higher commission error).

○ Presence
✕ Absence

Prob. Threshold = 0.03
Sensitivity = 1
(81 / 81)
Specificity = 0.74
(735 / 1000)
Kappa = 0.29

Figure 3. Example prediction maps illustrating the concepts of omission and commission error.

evaluation; this too is a large area of research which we do not summarize here. Fielding & Bell (1997) provide a widely-cited synopsis of model evaluation metrics. Here we focus on three metrics that are widely used to evaluate model performance, but these are by no means the only metrics that one could use. The first, and probably most widely-used performance metric, is the area under the receiver-operator characteristic curve, usually abbreviated as **AUC**. The AUC is a threshold-independent metric ranging from 0—1. High values of AUC indicate the tendency for predicted suitability at points known to be occupied by a species (the presences used in model validation) to be greater than predicted suitability at sites not known to be occupied by the species (absences or pseudo-absences from the validation data partition; Figure 4). In other words, a model will have a high AUC value if the predicted probabilities from the model at sites where the species is present (the presence data from the testing data subset) have, on average, higher values than sites where the species is 'absent' in the test data set.

The use of AUC has been criticized on numerous grounds (Lobo et al. 2007), the most problematic of which is the observation that 'good' AUC scores can be obtained simply by altering the geographic domain (a topic discussed later in the guidebook) for which models are created, making it easier for the model to more accurately differentiate presences and absences. Despite the criticisms, AUC remains in wide use for evaluating climate envelope models, at least in part because it is thought to be independent of prevalence (how rare or common a species is; Manel et al. 2001).

Among threshold-dependent performance metrics, **Cohen's kappa** (hereafter kappa) is arguably the most widely used. Kappa, like AUC, ranges from 0—1, with greater values again indicating greater performance. However, kappa measures the

classification ability of the model (using the 'test' data partition for model evaluation), or whether or not presences occur in grid cells categorized as suitable and absences occur in cells categorized as unsuitable. The model classification is generally summarized in the form of a **confusion matrix** (Table 2), which succinctly enumerates the number of correctly and incorrectly classified presences and absences.

Table 2. Confusion matrix summarizing correctly and incorrectly classified presences and absences. The boxes shaded in light grey are correctly classified, whereas the two dark grey boxes are incorrectly classified.

		Predicted	
		0	1
Observed			
0		998	2
1		15	60

A presence is considered correctly classified in the model if it occurs in a grid cell coded '1', indicating that the cell's predicted probability of climate suitability exceeds the threshold differentiating suitable and unsuitable areas. Conversely, a presence is incorrectly classified if it occurs in a grid cell coded '0', where the predicted suitability is below the threshold. Because some occurrences may be correctly classified by chance alone, the formulation of kappa attempts to correct for chance classification (see Franklin 2009 for more discussion on the issue and the formula for kappa). Like AUC, high values of kappa indicate greater classification ability. Unlike AUC, however, kappa is sensitive to the prevalence of species in a sample (e.g., the proportion of grid cells represented by species presences relative to the number of absences) and is not easily comparable among species that differ in prevalence. For that reason, there are not the same kinds of widely-accepted guidelines for interpreting 'good' versus 'poor' performance using kappa. Note that because AUC and kappa are measuring different things
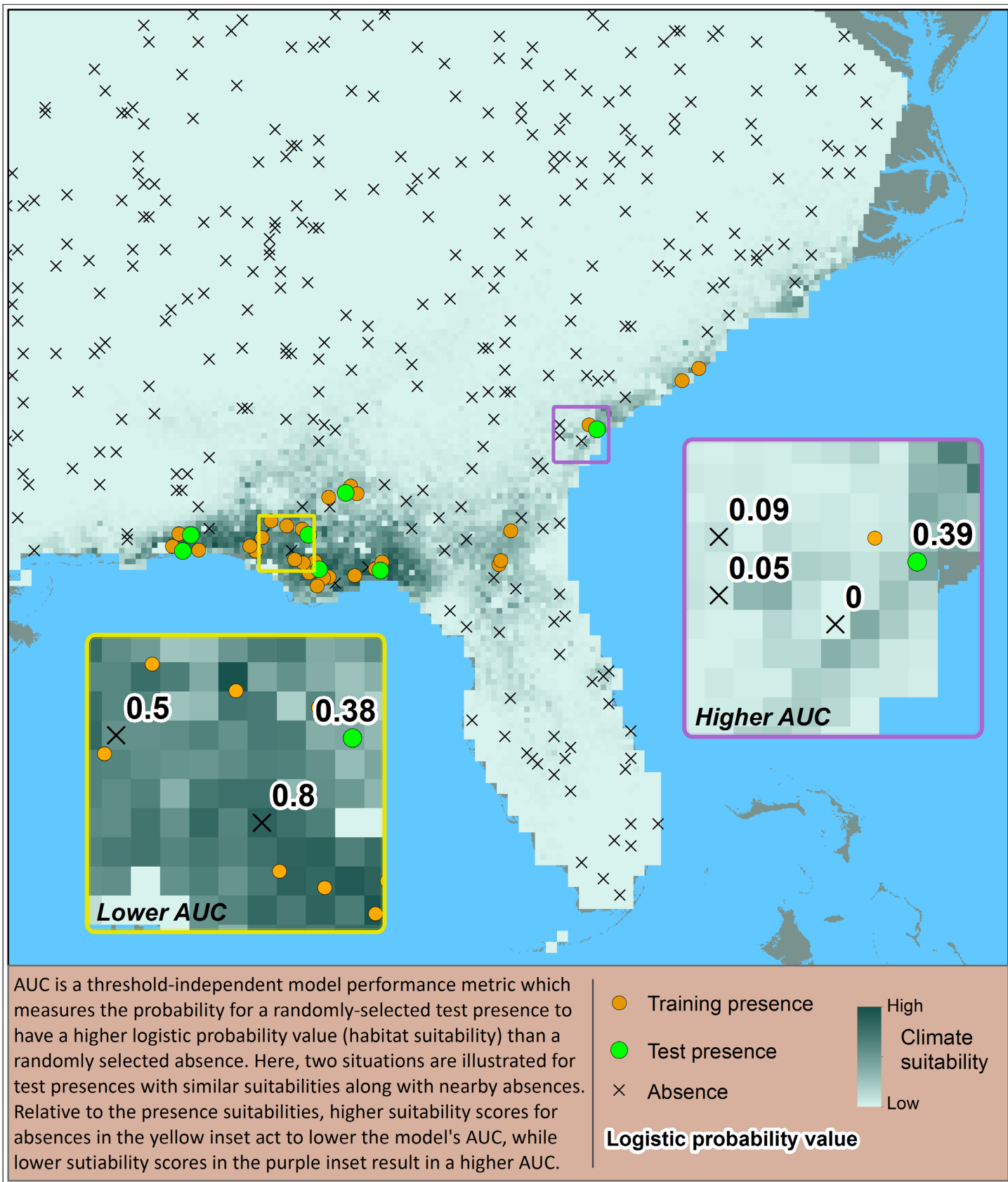
Figure 4. Example prediction maps illustrating the conditions that result in models with relatively high AUC scores (purple inset) and relatively low AUC scores (yellow inset).

about a model (AUC focuses on how well the model discriminates between occupied and unoccupied sites, where kappa focuses on classification ability), they are not always going to describe model performance similarly—it is not uncommon for models to have similar AUC scores, but very different kappa scores. Unfortunately there are no rules to resolve such apparent discrepancies, and users are encouraged to carefully consider what a model is being used for when interpreting evaluation metrics (e.g., will decision making require classifying suitable versus unsuitable areas—if so, kappa should probably take precedence over AUC in model evaluation), and ask for additional information when possible.

It is worth noting that many researchers consider metrics of model performance derived from the confusion matrix to be the most meaningful, because they are direct measurements of omission and commission error. From the confusion matrix, we may extract the percent of species presences that are correctly classified—these are 'true positives' and indicate the model's **sensitivity**. Because sensitivity measures the proportion of true presences that are correctly classified, it is directly related to omission error: (1-senstivity = omission rate). In other words, models with high sensitivity have low ommission error. We can also calculate the percent of absences that are correctly classified, the 'true negatives' that indicate model **specificity**. Specificity is directly related to commission error: (1-specificity = commission rate). In other words, models with high specificity have low commission error. In some cases, sensitivity and specificity are reported in their raw form as model diagnostics (Manel et al. 2001). It is also worth noting that sensitivity and specificity (or their linear transformations) lie at the heart of the calculation of AUC, although AUC is calculated across all possible thresholds rather than any particular threshold (hence the reference to AUC as a threshold-independent metric). See Franklin (2009) for more discussion on the issue.

A final performance metric that is less widely reported in the literature, but one that the authors of this guidebook use frequently is the spatial (or map) correlation between prediction maps (Syphard & Franklin 2009). **Map correlation** takes the suitability values from corresponding grid cells from two maps and calculates a correlation coefficient across all possible grid cells in two maps. The resulting metric describes the spatial correspondence between two maps; whereas the other metrics discussed focus primarily on classification, map correlation focuses on the spatial prediction maps themselves. The resulting correlation coefficient is interpreted just like Pearson's *r*, with values from 0—1, with higher values indicating greater correspondence between grid cell values in the two maps.

Suppose that two research groups are presenting results of their climate envelope models at a conference, or you are reading two papers by different authors. One presents their model evaluation criteria, stating that they infer excellent model performance because they report an AUC of 0.967. The other group had less apparent confidence in their model, because they report a kappa value of 0.742. Does that mean that the second group had a 'worse' model than the first one? Not at all—one is a threshold-dependent metric and the other not, so they are measuring different things about the model. It is entirely possible for two models to have similar AUC scores and very different kappas. Because of this ambiguity, it may be worth asking the two researchers after their talk or contacting them directly to describe their models sensitivity and specificity, which can be interpreted directly in terms of what matters most in model evaluation: commission and omission error.

How are continuous probabilities converted to presence-absence?

Sometimes when considering predictions from climate envelope models, we are interested in thinking about predictions of occurrence or climate suitability on a continuous scale from 0 (absent or completely unsuitable) to 1 (present or completely suitable). Sometimes, though, we want to say something more general about occupancy or suitability and identify areas where the species is expected (i.e., considered likely) to occur and areas where it is not. We want to draw a line in the sand and say 'we think the species will be here but not there'. This requires setting a threshold for the probability below which we will infer a species is likely to be absent, and above which we will infer presence. But how do we define this threshold? It would be very natural to say 'if the probability of presence is greater than 50%, we'll consider the species to be present, and if it's below 50%, we'll consider it to be absent'. However, it turns out that thresholds for determining presence and absence are highly associated with species prevalence (Manel et al. 2001). Therefore, threshold measures should generally try to take prevalence into account. One detailed study (Freeman & Moisen 2008) compared several ways to determine thresholds in climate envelope models and found ideal thresholds take prevalence into account—so if a species has low prevalence, a lower threshold is used to distinguish suitable and unsuitable areas. This topic is a complex one, and readers interested in the issue are referred to the papers by Freeman & Moisen (2008) and Manel et al. (2001). The bottom line, however, is that using an arbitrary 50% threshold to differentiate suitable and unsuitable areas in a map is often not defensible in terms of ecological relevance.

What algorithms are used to model the species-climate relationship?

The choice of modeling algorithm is probably the single most important choice a user makes when creating a climate envelope model. Several studies indicate that performance varies more as a function of the algorithm used to create a model than other components of the modeling process (e.g., the number of occurrence points, predictor variables, or climate dataset; Watling et al. 2012; Elith & Graham 2009; Dormann et al. 2008; Elith et al. 2006). That being said, studies do not always agree as to which algorithm is the best; regression trees, for example, have been implicated as both high-performing and low-performing across different studies (Elith & Graham 2009 ; Meynard & Quinn 2007). Many authors have suggested that users create several models using different algorithms, and average the observations across those various models (Araújo & New 2007).

It is beyond the scope of the current work to review all of the algorithms available for climate envelope modeling. Rather than review all the algorithms available, we present data on model performance and spatial predictions for a subset of modeling algorithms that are widely used by many researchers for climate envelope modeling. We focus on the implications of different algorithms for modeling rather than a detailed explanation of the theory underlying algorithms or the pros and cons of alternative algorithms.

We compare performance of climate envelope models using three modeling algorithms: **generalized linear models** (GLM; McCullugh & Nelder 1989), **maximum entropy** models (Max; Phillips et al. 2006; Phillips & Dudík 2006) and **random forests** (RF; Cutler et al. 2007). All three of these methods

are widely used in climate envelope modeling, and we have used them for our work on threatened and endangered vertebrates in Florida. Briefly, the GLM approach is essentially a logistic regression in which presence and (pseudo-) absence are modeled as a binomial response against climate predictors. The maximum entropy model uses knowledge of probability distributions of climate variables at sites known to be occupied by the species as well as a random sample of background climate conditions to estimate the probability distribution of species occurrence. Random forests are a classification approach in which many random subsets of climate predictors are used to classify presence and (pseudo-) absence repeatedly in order to find values of climate predictors that best classify presence and absence.

We describe performance using two of the model evaluation metrics described previously, AUC and Cohen's kappa. Although there were not huge differences in AUC among the three modeling algorithms (Table 3), differences become more pronounced using Cohen's kappa, which describes a model's ability to accurately classify presences and pseudo-absences. On average, kappa was lowest for GLMs, intermediate for maximum entropy models, and highest using random forests.

Suppose you hear about work from two research groups doing climate envelope modeling, but one uses generalized linear models, and the other group is using maximum entropy. Both describe their model evaluation procedure and each independently concludes that, based on exceptionally high AUC scores (> 0.950 in both cases), they have high-performing models. Does that mean that prediction maps resulting from their work will be congruent? In other words, do two different models with 'high' performance result in prediction maps that look the same? Not necessarily. To illustrate the point, we

include two prediction maps for one of the species the authors of this guidebook are working with, the Florida sand skink (Figure 5). This endangered lizard is endemic to highly fragmented habitat patches distributed in a narrow band through central peninsular Florida. Knowing what the geographic distribution of the Florida sand skink looks like, it is easy to see that the random forest model does a better job of delimiting a climate envelope for this species than the generalized linear model. But imagine if the known distribution of a species were not as well-characterized? Which map would we believe? Without additional information, we would not know which model was 'right', which is why we suggest that results of multiple models be used together to form more robust conclusions. Remember that modeling

Table 3. Variation in model performance among three algorithms used for climate envelope modeling

	AUC			Kappa		
	GLM	Max	RF	GLM	Max	RF
Common name						
Mammals						
Florida panther	0.981	0.995	0.990	0.262	0.668	0.890
Birds						
Florida grasshopper sparrow	0.987	0.998	0.983	0.065	0.310	0.773
Florida scrub jay	0.998	0.990	0.999	0.718	0.526	0.902
Piping plover	0.865	0.927	0.989	0.168	0.378	0.836
Wood stork	0.910	0.947	0.980	0.422	0.507	0.863
Audubon's crested caracara	0.957	0.993	0.999	0.114	0.712	0.897
Everglade snail kite	0.994	0.995	0.999	0.322	0.540	0.663
Whooping crane	0.918	0.986	0.989	0.102	0.351	0.673
Red-cockaded woodpecker	0.964	0.958	0.981	0.273	0.341	0.707
Reptiles and amphibians						
American crocodile	0.931	0.942	0.979	0.078	0.126	0.522
Sand skink	0.998	0.996	0.997	0.234	0.111	0.734
Eastern indigo snake	0.979	0.996	0.999	0.354	0.552	0.895

algorithms really matter and can sometimes make very different spatial predictions, even when everything else about the model is the same.

What are some of the assumptions underlying climate envelope modeling?

When considering what climate envelope models do and don't do, it is worth keeping in mind some of the assumptions underlying the models (see Wiens et al. 2009 for a review). One key assumption is

Neoseps (Plestiodon) reynoldsi (Florida sand skink)

RF

GLM

Prediction maps for the Florida sand skink made using the same set of occurrence and climate data but with two different modeling algorithms, Random Forests (RF) and Generalized Linear Model (GLM). Both algorithms showed good performance according to the AUC metric, but their spatial predictions differ considerably. The sand skink's current range is outlined in black.
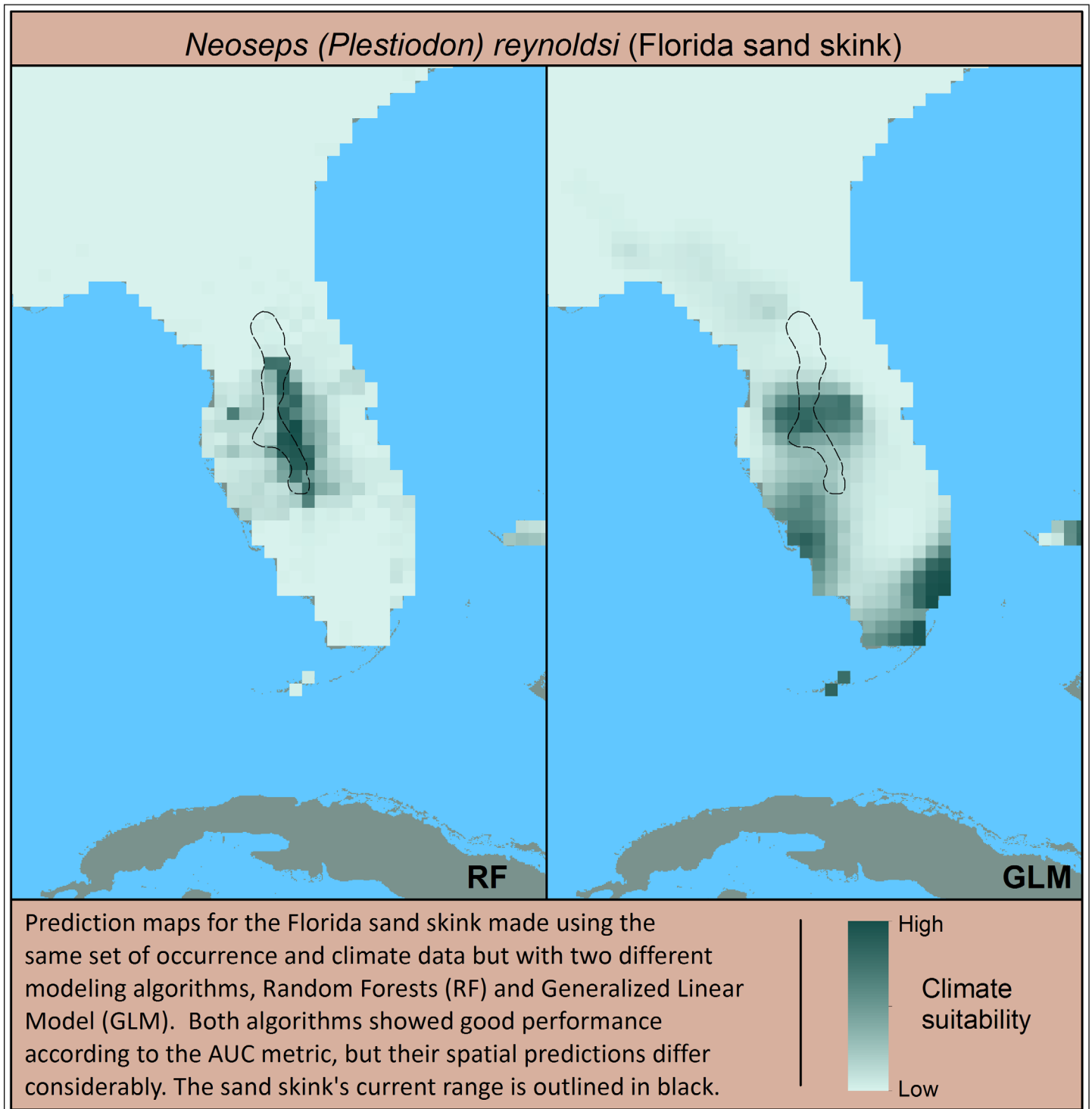
High

Climate suitability

Low

Figure 5. Prediction maps for the Florida sand skink illustrating different predictions resulting from the use of two different modeling algorithms.

that of environmental equilibrium, or the idea that species occur throughout the range of suitable areas available to them. By including species occurrences in models, we are assuming that those occurrences represent persistent populations with non-zero and non-negative growth rates. A lot of work in ecology has investigated **source-sink population dynamics**, and it is widely appreciated that species can occur in low-quality 'sink' areas that would not, in and of themselves, support viable populations in the absence of dispersal from 'source' populations (Pulliam 1988). Incorporating sink populations, or observations of vagrant individuals passing through an area may result in overpredicting a species climate envelope by including areas not suitable for the long-term persistence of the species. Conversely, species may be artificially absent from areas that are environmentally suitable. Violation of the assumption that species occurrences represent the full breadth of environmental conditions suitable for the species can lead to models that under or over-represent the climate envelope of a species, resulting in models with high omission or commission errors, respectively.

Another important assumption to bear in mind when using climate envelope models to project climate change effects on species distributions is that of **niche conservatism**. By defining a species-climate relationship on the basis of contemporary climate and extrapolating that relationship into the future using projected climate data, we assume that a species will not evolve the ability to adapt to new climates. Just as failing to consider species occurrences from areas of suitable climate conditions can result in models that underpredict the distribution of climate suitability, violation of the assumption of niche conservatism (no adaptation) can result in models that underpredict areas of future climate suitability. There are many examples of species that show adaptations to local climate conditions (i.e., herbs, Souther & McGraw et

al 2011,;;trees, Vitasse et al. 2009; sea turtles, Weber et al. 2012), suggesting that many species have the potential for *in situ* adaptation in the face of climate change. Climate envelope models will likely make poor predictions when modeling species that are capable of rapidly adapting to new climate conditions outside the range of conditions experienced elsewhere in their geographic range.

For more information:

Dormann C. F., O. Purschke, J. R. García Márquez, S. Lautenbach & B. Schröder. 2008. Components of uncertainty in species distribution analysis: a case study of the Great Grey Shrike. Ecology 89:3371—3386.

Elith J., M. A. Burgman & H. M. Regan. 2002. Mapping epistemic uncertainties and vague concepts in predictions of species distribution. Ecological Modelling 157:313—329.

Elith, J., C. H. Graham, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. McC. Overton, A. Townsend Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberón, S. Williams, M. S. Wisz & N. E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129—151.

Elith J., S. J. Phillips, T. Hastie, M. Dudík, Y. E. Chee & C. J. Yates. 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17:43—57.

Fielding A. H. & J. F. Bell. 1997. A review of

methods for the assessment of prediction errors in   conservation presence/absence models. Environmental Conservation 24:38—49.

Hernandez P. A., C. H. Graham, L. L. Master & D. L. Albert. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29:773—785.

Meynard C. N. & J. F. Quinn. 2007. Predicting species distributions: a critical comparison of the most common statistical models using artificial species. Journal of Biogeography 34:1455—1469.

Souther S. & J. B. McGraw. 2011. Evidence of local adaptation in the demographic response of American ginseng to interannual temperature variation. Conservation Biology 25:922—931.

Vitasse Y., S. Delzon, C. C. Bresson, R. Michalet & A. Kremer. 2009. Altitudinal differentiation in growth and phenology among populations of temperate-zone tree species growing in a common garden. Canadian Journal of Forest Research 39:1259—1269.

Weber S. B., A. C. Broderick, T. G. G. Groothuis, J. Ellick, B. J. Godley & J. D. Blout. 2012. Fine-scale thermal adaptation in a green turtle nesting population. Proceedings of the Royal Society B-Biological Sciences 279:1077—1084.

Among the most obvious components of a climate envelope model are the inputs used to construct the model, namely the georeferenced species occurrences and the climate data. Here we treat issues related to each of these modeling components separately, beginning with the species occurrence data, followed by the climate data layers.

The occurrence data used for climate envelope modeling can come from any number of sources, including field observations, radio tracking studies, or the primary literature. Many occurrence data are compiled in online databases such as the Global Biodiversity Information Facility (www.gbif.org) (information on additional online sources of occurrence data are included at the end of the guidebook). Although online databases are a convenient source of occurrence data, one should be prepared to critique input data as part of the overall interpretation process. In this section, we will introduce some critical considerations about data validity, format, and integrity. Errors in georeferencing can occur, and sometimes different data types are mixed (such as inclusion of zoo or fossil records along with observations of 'wild' individuals). It can take much longer to prepare the species occurrence data for modeling (correcting locational errors when possible, removing duplicate observations, etc) than it takes to do the modeling itself.

Although we do not go into detail here, a large literature deals with differentiating apparent absence (the non-detection of a species) from true absence (a species really does not occur in an area) in biodiversity surveys. Unfortunately, systematic surveys are often not available for most species, so model developers rely on more-or-less haphazardly collected data consisting primarily of opportunistic species encounters. We call these data 'presence only' because they do not include true survey-based absences.

Some of the algorithms used to create climate envelope models are specifically designed for use with presence only data ('presence only methods') whereas other methods are designed for use with both presence and absence data. There are also methods available for researchers who want to use presence-absence methods even if they do not have access to true survey-based absence data. Both of these topics are covered in more detail in other sections of the guidebook.

Species Occurrences

What is the effect of erroneous occurrence data on models?

The georeferenced occurrences are the only species-specific data that are fed into a climate envelope model, so care should be taken to ensure that errors in the occurrence data are minimized. Published studies have investigated the effect of incorrect occurrence data by intentionally manipulating species occurrences to introduce error (Guisan et al. 2007). Those results suggest that errors in the coordinate data can decrease model performance across a range of algorithms, although the observed decrease was not as substantial as may be expected (e.g., a decrease in AUC from about

0.80 for the best-performing algorithms using unmanipulated data to about 0.75 for models in which error was artificially introduced). It can be hard to know when an occurrence point is an error or not. Some locational errors are obvious; the author's of this guidebook have found that data obtained from online databases sometimes contain equivocal information, the most common of which is a latitude or longitude coordinate missing a negative sign (-) for values south of the equator or west of the prime meridian. In many cases we have found sufficient supplementary information (e.g., a place name) accompanying the coordinate data so that it was possible to determine the nature of the error and correct it. In other cases, we have found suspect outlying occurrences thousands of kilometers from the nearest georeferenced observation with no obvious explanation. Recall that one of the assumptions in climate envelope modeling is that each observation represents an area where environmental conditions (climate) are suitable for the maintenance of a viable population (e.g., one with a non-negative growth rate over time) of the target species. If an observation is from an area where climate is not suitable for long-term population persistence, it should probably be excluded from analysis. This same criterion can be used to eliminate other types of occurrences that may not necessarily represent error per se. For example, we have also observed that some occurrences compiled in online databases include zoo populations or fossil records. Since those observations obviously represent populations that are not at equilibrium with contemporary environmental conditions, they should also be excluded from analysis. Other sources of error in the species occurrence data may be more difficult to detect. In our own work on vertebrate species in Florida, we elected to include most data unless there was a very obvious error in the identity of coordinates or observations were known to represent non-viable populations.

What about the effects of small sample size?

Another issue of concern with the occurrence data is the number of observations. As with the effects of erroneous location data, researchers have investigated the effect of sample size on model performance. Available data suggests that small sample sizes are not necessarily associated with decreased model performance, especially when sample size is greater than about 20—30 occupied grid cells (Guisan et al. 2007). In Table 4 we have added a column describing sample size of the species occurrence data for models included in the comparison of algorithms earlier in the guidebook. We focus specifically on AUC and Cohen's kappa for models using the random forest algorithm. Inspection of the table indicates that performance is independent of sample size.

Table 4. Model performance metrics and sample size for climate envelope models for threatened and endangered vertebrates in Florida using the random forest algorithm.

	Number of occurrence points	AUC	Kappa
Common name			
Mammals			
Florida panther	86	0.990	0.890
Birds			
Florida grasshopper sparrow	26	0.983	0.773
Florida scrub jay	194	0.999	0.902
Piping plover	782	0.989	0.836
Wood stork	1435	0.980	0.863
Audubon's crested caracara	159	0.999	0.897
Everglade snail kite	101	0.999	0.663
Whooping crane	176	0.989	0.673
Red-cockaded woodpecker	515	0.981	0.707
Reptiles and amphibians			
American crocodile	116	0.979	0.522
Sand skink	18	0.997	0.734
Eastern indigo snake	278	0.999	0.895

More important than sample size per se is the extent to which the full range of climate conditions experienced by a species is included in the occurrence datasets. If no occurrence data are available for areas occupied by the species where climate differs from other occupied areas, the model will probably not identify those areas as suitable for the species. Thus, we try to make sure that observations represent the full range of conditions experienced by the species, rather than ensuring representation of all possible species observations.

Prior to analysis, duplicate occurrences within a grid cell are removed, so many individual occurrences ultimately are not included in analysis anyway. The bottom line is that many studies demonstrate that predictions from the best-performing algorithms are robust to small sample sizes (Wisz et al. 2008), so having few occurrence records should not be a deterrent to modeling *as long as those occurrences represent the full range of climate conditions experienced by the species*.

How are pseudo-absence data obtained?

Because we often lack true survey-based absence data when constructing climate envelope models, it is common to use randomly-selected '**pseudo-absences**' (Chefaoui & Lobo 2008) along with known species presences. Pseudo-absences are really a random sample of 'background' environmental conditions, rather than a description of conditions at places where a species is actually known to be absent. There are different ways to select pseudo-absences, but in general a random sample of points (often between 1,000-10,000) is selected from the study area, and climate data are extracted for this random sample. Although it seems like a relatively straightforward issue, the selection of pseudo-

absences can have a substantial influence on model performance and predictions. This is because the selection of pseudo-absences defines the modeling domain, or the geographic area for which a model is being constructed. It's very easy to think about constructing a climate envelope model for a given area, probably politically defined (e.g., the continental USA, California, Spain) and moving forward from there. The problem is that species, and the climate (and other) conditions that limit where species occur, do not necessarily coincide with political boundaries. A not unrelated issue has to do with potential biases in the species occurrence data themselves. It is not uncommon, especially in more remote areas, that opportunistic species observations are made in relatively more accessible areas that may have different environmental conditions than the region as a whole (think about species observations concentrated along roads through a densely forested area, for example). It has been suggested that one way to deal with the fact that species presences themselves may be biased as well as reduce some of the arbitrariness of defining the modeling domain is to use what is called the **target group approach** (Phillips et al. 2009). But before introducing the target group approach, let's focus for a minute on the general problem: how does the selection of pseudo-absences influence model performance and predictions?

Researchers have looked at the question of how the area from which pseudo-absences are drawn affects model outcomes by doing experiments where pseudo-absences are drawn from an increasing distance from where presences are concentrated (VanDerWal et al. 2009). As the size of the pseudo-absence 'background' increases, AUC tends to increase as well. Simply put, it becomes easier and easier for a model to differentiate suitable and unsuitable areas when the background includes lots of unsuitable areas. This makes sense if you think about it at a really

large scale: it's easy to differentiate the climate in say, northern California, from the climate in the rest of the continental USA. But it's probably a lot harder to differentiate the climate in northern California from the climate in the rest of the Pacific Northwest. The result is that model performance statistics like AUC get larger as the background area gets bigger (at some point, the background can get too big and performance starts to decrease again). So a researcher can more or less guarantee a high AUC by varying the model domain appropriately. Hence the need to come up with a non-arbitrary way to define model domain.

The target group approach has been recommended as one way to reduce some of the arbitrariness in defining model domain, while also dealing with another problem: that of bias in the presence data themselves. The idea is that if presences are taken from a restricted portion of the environment (e.g., roads) that differs systematically from areas without roads (and hence no species observations), you may violate one of the assumptions of the models, namely that presences are obtained from all suitable areas where the species occurs. Phillips et al. (2009) suggest that the best way to fight bias is with bias: rather than using pseudo-absences from random points, take observations of ecologically similar species sampled using similar methods (therefore incorporating the same sample selection bias as in the presence data), and use those observations (where similar species are known to occur, but the study species has not been recorded) as the pseudo-absences in the focal species model. Although Phillips et al. (2009) suggest using the individual observations of closely related species as the pseudo-absences for the species being modeled, the authors of this guidebook have used a modification of the target group approach. Under their approach, a polygon is drawn around the observations of target group species to define an area from which pseudo-absence data are randomly selected (Figure 6). This approach reduces some of the arbitrariness of defining the model domain, but does not rely on a potentially small number of target group occurrences for modeling.
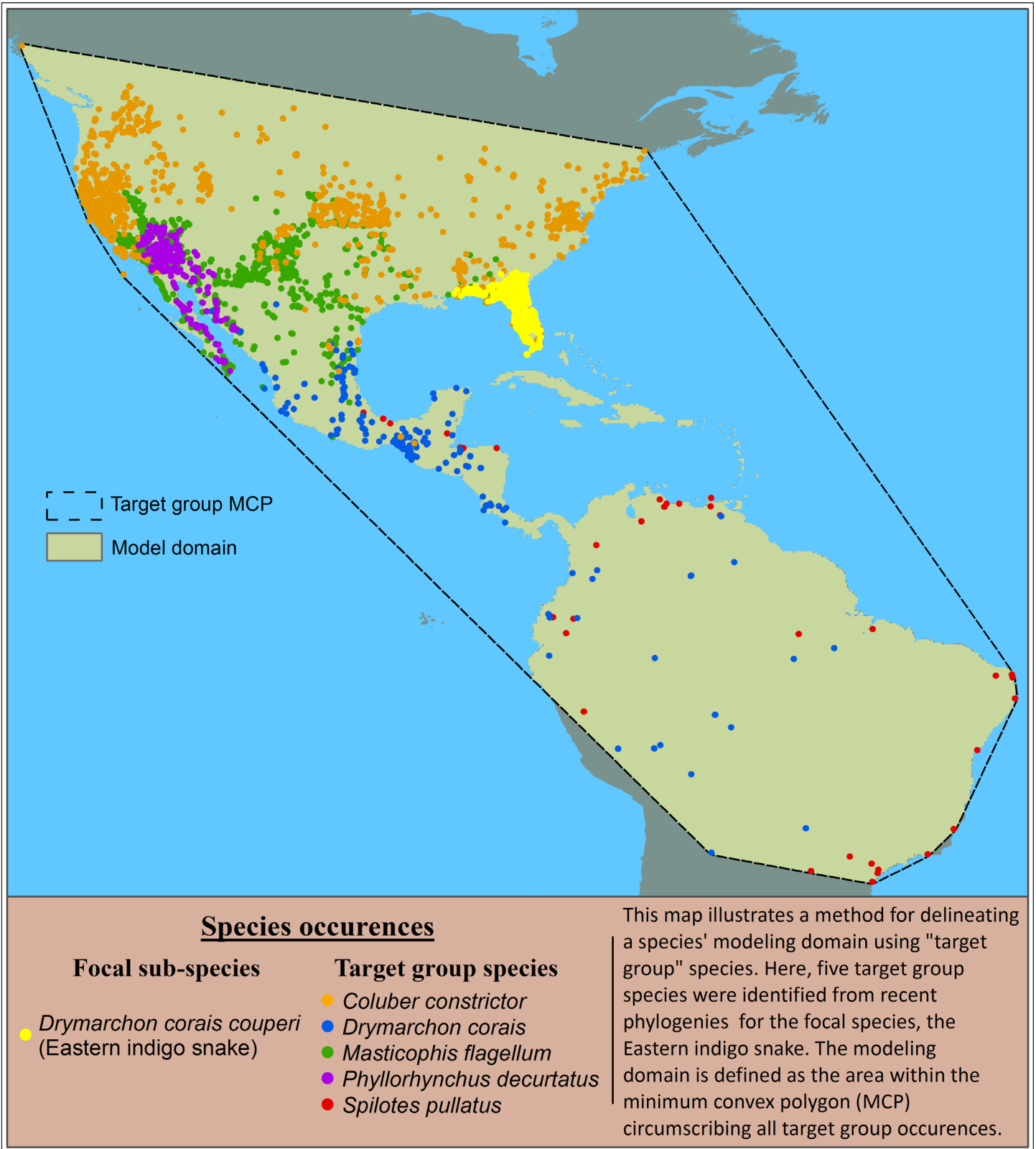
Climate Data

In addition to the georeferenced species observations, the other primary data input in climate envelope models are the climate data themselves. Climate data are generally used in the form of raster grids. These grids are used to sample climate from grid cells occupied by a species, and are also used to describe climate in areas to which the model will be extrapolated. Here we focus almost exclusively on the contemporary climate data used for model calibration rather than the many issues associated with describing future climate.

Which climate data should be used for models?

The selection of a climate data set for modeling of the contemporary species-climate relationship should be made in light of many considerations: the spatial domain of the study, grid cell size, availability of relevant data layers, etc. For example, several high-resolution climate datasets exist for some or all of the United States, but these datasets are not relevant for modeling species that occur outside of the United States. Similarly, some datasets may have global extent, but only for a limited number of climate variables (e.g., monthly mean temperature and monthly precipitation, but not minimum or maximum temperature). Grid cell size also matters: when modeling a range-restricted species, the entire geographic range may conceivably fall within a single grid cell measuring 1/6 of a decimal degree in each direction. Thus you may have sufficient species

Species occurences

Focal sub-species

- ● *Drymarchon corais couperi* (Eastern indigo snake)

Target group species

- ● *Coluber constrictor*
- ● *Drymarchon corais*
- ● *Masticophis flagellum*
- ● *Phyllorhynchus decurtatus*
- ● *Spilotes pullatus*

This map illustrates a method for delineating a species' modeling domain using "target group" species. Here, five target group species were identified from recent phylogenies for the focal species, the Eastern indigo snake. The modeling domain is defined as the area within the minimum convex polygon (MCP) circumscribing all target group occurences.

Figure 6. Illustration of the authors' modification of the target group approach (Phillips et al. 2009).

observations (say 100+) for modeling, but if those occurrences fall within a single grid cell, you have a sample size of one. In that case, it may be necessary to use a higher-resolution climate dataset so that more occurrences can be effectively incorporated into models.

There are many sources of data describing contemporary climate that can be used as inputs to climate envelope models. Although we cannot mention all possible data sources here, we mention some of the most widely-used and freely-available climate datasets. The PRISM dataset includes climate normals (long-term averages) for the 1971—2000 and 1981—2010 periods for the contiguous United States at a resolution of 30 arc-seconds (1/120th of a decimal degree). In addition, monthly climate data are available for the period 1895—present. The monthly data are available at a resolution of 2.5 arc-minutes (1/24th of a decimal degree). Data from the Lawrence Livermore National Laboratory (LLNL) are available as climate normals for the years 1950—1999 at a 1/8 degree spatial resolution for the contiguous United States. Summary information on these and other sources of spatially-explicit climate maps are included in Table 5.

Although we do not go into detail on the many alternative general circulation models (GCMs) used to forecast climate change effects, note that these and other issues such as the considerations involving the selection of emissions scenarios and the use of downscaled data have been addressed in other documents (e.g., Glick et al. 2011).

Which individual variables should be used for modeling?

There are two broad types of temperature and precipitation data that can be used for climate envelope modeling: data can be expressed as monthly variables (e.g., January mean temperature, July precipitation) or as bioclimate variables (Nix 1986) that largely describe seasonal trends and climate extremes (the full list of bioclimate variables is available in Appendix I). Because detailed species-specific data are not incorporated into climate envelope models, variable selection is the only way to 'tailor' models to species ecology. Many authors have advised careful attention to variable selection so as to include variables that have direct and proximal effects on species ecology (Austin 2002, 2007). That is, variables should have a known direct influence or

Table 5. Summary information for several contemporary climate datasets useful for climate envelope modeling.

Dataset	Website	Domain	Variables	Time period covered	Resolution
WorldClim	http://www.worldclim.org	Global	Min T, Max T, Mean T, Precipitation, Bioclim	1950—2000	30 arc-seconds – 10 arc-minutes
Climate Research Unit	http://www.cru.uea.ac.uk	Global	Mean T, Diurnal T range, Precipitation, Ground-frost frequency,	1961—1990	10 arc-minutes
PRISM	http://www.prism.oregonstate.edu	48 conterminous United States	Min T, Max T, Mean T, Precipitation, Average dewpoint	1896—2011	2.5 arc-minutes
LLNL	http://gdo-dcp.ucllnl.org	48 conterminous United States	Precipitation (mm/day), T mean	1950—1999	2 degrees
NCEP	http://dss.ucar.edu/datasets/ds093.2	Global	Min T, Max T, Mean T, Precipitation, others	1979—2011	30 arc-minutes—2.5 degrees

be considered likely to have a direct physiological influence on the species of interest (direct) and be the most causal variables describing a physiological response (proximal). For example, temperature is generally a more proximal variable than elevation, because species do not respond to elevation directly, but more likely respond to temperature (which of course varies as a function of elevation). For a species distributed in relatively warm areas, winter minimum temperature is probably among the most direct variables to include in a model, because there is a clear association between the variable (winter temperature) and species physiology (warm-climate species are generally freeze intolerant). The feeling is that in the absence of judicious variable selection, model predictions will be arbitrary rather than reflective of 'true' responses of species to changing climate.
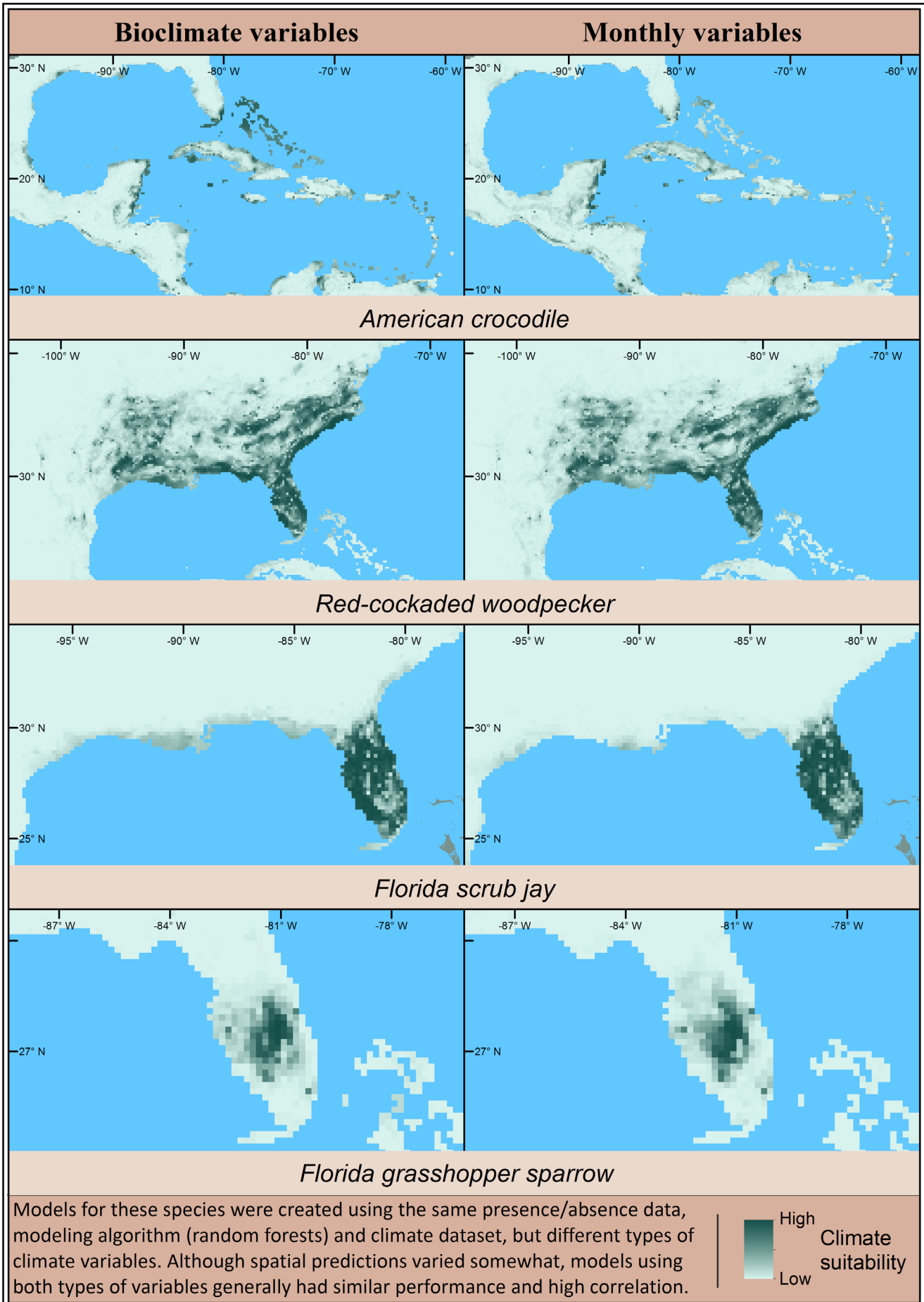
Because the authors of this guidebook were unaware of studies comparing models constructed from monthly climate variables to those constructed from bioclimate variables, they conducted such an investigation with some of the species from their work on threatened and endangered vertebrates in Florida (Watling et al. 2012). They found that model performance (using AUC and kappa) was similar, and spatial correlations very high for models using bioclimate and monthly variables with the random forest algorithm (Table 6, Figure 7; models using GLMs generally had lower performance and less correlated spatial predictions than random forest models).

An issue worth considering when selecting predictor variables for use in climate envelope models is data collinearity. **Collinearity** refers to a situation in which multiple predictor variables are highly correlated with one another, undermining the ability to determine which predictor is truly driving a response, and to what degree. It has been suggested that collinearity can affect predictions from climate envelope models, and researchers often recommended that highly intercorrelated variables be removed from analysis (Dormann et al. 2008). There are no hard-and-fast rules for determining what constitutes intercorrelated versus uncorrelated variables, and this decision is usually made operationally by individual researchers. It's also worth mentioning that in many cases, there is no mention of data collinearity when discussing models. As a general rule, though, model results are probably going to be more interpretable if they use fewer, relatively uncorrelated variables; in that case it should be more straightforward to interpret shifts in climate suitability unambiguously in terms of the climate predictors.

Table 6. Spatial correlations between prediction maps from models created using bioclimate or monthly climate variables. The overall high correlations indicate that prediction maps were similar regardless of the identity of variables used.

Species	Spatial correlation
Mammals	
Florida panther	0.867
Birds	
Florida grasshopper sparrow	0.946
Florida scrub jay	0.968
Piping plover	0.903
Wood stork	0.882
Audubon's crested caracara	0.953
Everglade snail kite	0.932
Whooping crane	0.912
Red-cockaded woodpecker	0.960
Amphibians and Reptiles	
American crocodile	0.671
Sand skink	0.921
Eastern indigo snake	0.970
Average ± 1 SD	0.907 ± 0.081

Bioclimate variables | **Monthly variables**

American crocodile

Red-cockaded woodpecker

Florida scrub jay

Florida grasshopper sparrow

Models for these species were created using the same presence/absence data, modeling algorithm (random forests) and climate dataset, but different types of climate variables. Although spatial predictions varied somewhat, models using both types of variables generally had similar performance and high correlation.

Climate suitability: High — Low

Figure 7. Example prediction maps showing similarities in predictions from climate envelope models constructed from bioclimate and monthly climate variables.

The extent to which climate variables co-vary depends on both variable identity and spatial scale. The guidebook authors compared global data on monthly temperature and precipitation and found that temperature variables tend to show greater collinearity than precipitation variables (Table 7). In fact, only two monthly temperature variables are relatively uncorrelated at the global scale: January and July. Temperature during other months of the year tend to be highly correlated ($r > 0.85$) with either January or July temperature. Precipitation variables tend to be less intercorrelated at a global scale. Relatively independent ($r < 0.85$) precipitation variables include January, April, May, June, September and October (Table 7). At smaller spatial scales, patterns of collinearity may depart from global averages. In general, one may expect to see greater collinearity among variables at smaller spatial scales, although there may be considerable spatial variation in the extent to which this is true. For example, in areas of rapid elevation change or varied topography, climate may covary much less strongly than in relatively flat areas.

Although it is increasingly common to see the issue of data collinearity dealt with explicitly when presenting results from climate envelope models, some researchers still present results where variable

Table 7. Correlation coefficients between pairwise combinations of monthly mean temperature (top) and precipitation (bottom) in the WorldClim global climate dataset based on records from approximately 1950—2000.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Jan	1	0.997	0.979	0.933	0.852	0.696	0.585	0.690	0.847	0.945	0.987	0.998
Feb	0.997	1	0.990	0.953	0.879	0.731	0.625	0.725	0.872	0.960	0.991	0.997
Mar	0.979	0.990	1	0.984	0.930	0.804	0.708	0.795	0.919	0.982	0.990	0.982
Apr	0.933	0.953	0.984	1	0.978	0.887	0.806	0.874	0.961	0.987	0.966	0.941
May	0.852	0.879	0.930	0.978	1	0.961	0.901	0.943	0.980	0.960	0.907	0.865
Jun	0.696	0.731	0.804	0.887	0.961	1	0.982	0.987	0.955	0.871	0.777	0.715
Jul	0.585	0.625	0.708	0.806	0.901	0.982	1	0.987	0.916	0.798	0.683	0.610
Aug	0.690	0.725	0.795	0.874	0.943	0.987	0.987	1	0.963	0.874	0.777	0.712
Sep	0.847	0.872	0.919	0.961	0.980	0.955	0.916	0.963	1	0.969	0.908	0.863
Oct	0.945	0.960	0.982	0.987	0.960	0.871	0.798	0.874	0.969	1	0.980	0.955
Nov	0.987	0.991	0.990	0.966	0.907	0.777	0.683	0.777	0.908	0.980	1	0.993
Dec	0.998	0.997	0.982	0.941	0.865	0.715	0.610	0.712	0.863	0.955	0.993	1

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Jan	1	0.981	0.926	0.729	0.440	0.163	0.035	0.028	0.218	0.543	0.805	0.956
Feb	0.981	1	0.960	0.777	0.477	0.185	0.048	0.038	0.224	0.532	0.774	0.922
Mar	0.926	0.960	1	0.887	0.599	0.282	0.129	0.119	0.315	0.604	0.798	0.888
Apr	0.729	0.777	0.887	1	0.844	0.535	0.351	0.338	0.516	0.720	0.775	0.736
May	0.440	0.477	0.599	0.844	1	0.846	0.657	0.641	0.743	0.752	0.604	0.474
Jun	0.163	0.185	0.282	0.535	0.846	1	0.911	0.875	0.847	0.634	0.360	0.205
Jul	0.035	0.048	0.129	0.351	0.657	0.911	1	0.96	0.839	0.522	0.222	0.077
Aug	0.028	0.038	0.119	0.338	0.641	0.875	0.960	1	0.897	0.574	0.248	0.082
Sep	0.218	0.224	0.315	0.516	0.743	0.847	0.839	0.897	1	0.818	0.498	0.301
Oct	0.543	0.532	0.604	0.720	0.752	0.634	0.522	0.574	0.818	1	0.851	0.654
Nov	0.805	0.774	0.798	0.775	0.604	0.360	0.222	0.248	0.498	0.851	1	0.911
Dec	0.956	0.922	0.888	0.736	0.474	0.205	0.077	0.082	0.301	0.654	0.911	1

collinearity is an issue. If you suspect that models you are hearing or reading about include highly intercorrelated variables, it is worth asking whether the authors tested for collinearity—if not, treat the results with extra caution. Including highly correlated variables in an analysis does not necessarily always change the model outcomes in a major way, but it can influence spatial predictions from climate envelope models.

How does the climate dataset input influence model outputs?

Although this issue has received surprisingly little attention in the vast literature on species distribution modeling, there is evidence that alternative climate data inputs can under some circumstances affect the spatial predictions of climate envelope models (Parra & Monahan 2008). Authors of this guidebook have investigated differences in performance and spatial predictions between climate envelope models constructed with data from two global climate datasets, WorldClim (Hijmans et al. 2005) and the Climate Research Unit (CRU; New et al. 2002); summary data on both datasets are included in Table 5. Both datasets are widely-used for a variety of climate change applications, and represent average climate conditions (temperature and precipitation) for the terrestrial portion of the globe from the mid to late twentieth century. Both datasets were created by spatially interpolating data derived from long-term records from a global network of weather stations, although the identity and number of stations, the elevation model used as a covariate to climate, and the spatial interpolation technique used to create the climate surface varied between datasets (Hijmans et al. 2005). We extracted the same subset of monthly temperature and precipitation variables from each dataset (the identity of variables differed for each species and represented combinations with high

explanatory power and relatively low covariance with other variables) and constructed separate models for the same group of 12 threatened and endangered vertebrate species in Florida. We looked at model performance using AUC and kappa, and used map correlation to describe similarities between the two maps for each species. Although we found evidence that model performance varied among algorithms (which is not a surprise given that many studies have shown that algorithm selection is the primary determinant of model performance), there were no significant differences in AUC or kappa between models created with CRU or WorldClim data. Average spatial correlations between prediction maps made with CRU or WorldClim data ranged from 0.817—0.886 depending on the algorithm used. By way of comparison, recall that we found higher spatial correlations, and more consistent predictions (e.g., there was a lower standard deviation) when comparing bioclimate versus monthly variables. To put these numbers in context, we also looked at spatial correlations between projections of future climate using three different GCMs. The average spatial correlations among GCMs ranged from 0.763—0.819), only a bit lower than average correlations between contemporary climate datasets. So while differences between the contemporary climate data incorporated into models does not have as strong an effect, on average, as the selection of modeling algorithm, it can introduce variation in spatial predictions made from models. Different researchers use lots of different sources of contemporary climate data for climate envelope modeling (Table 5 is just a summary of a few datasets, and is far from an exhaustive list), and rarely describe how different their results are when using inputs from different datasets. If you are seeing unexplained discrepancies between maps that are otherwise similar (i.e., made for the same species in the same area using the same modeling algorithm), it is worth keeping in mind that

some differences between prediction maps may be expected just because of the different data sets used to create the models—ask the researchers to describe their input climate data to determine whether this can explain differences in model predictions.

Considering the effects of climate dataset in the context of results presented at a climate change conference or in two different journal articles, bear in mind that the algorithm used for modeling is most likely to be the biggest source of variation when comparing results. Comparing results generated using different algorithms can be like comparing apples and oranges. Results obtained using different climate datasets are probably more comparable, on average, than results using different algorithms, but when confronted with differing model predictions, it is worth noting that some results could be a function of differences in how climate is described in two different datasets.

For more information:

Hijmans R. S., S. E. Cameron, J. L. Parra, P. G. Jones & A. Jarvis. 2005. Very high resolution climate surfaces for global land areas. International Journal of Climatology 25:1965—1978.

New M., D. Lister, M. Hulme & I. Makin. 2002. A high-resolution data set of surface climate over global land areas. Climate Research 21: 1—25.

Phillips S. J., M. Dudík, J. Elith, C. M. Graham, A. Lehmann, J. Leathwick & S. Ferrier. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological Applications 19:181—197.

To what extent can model results be extrapolated to other contexts?

There has been a lot of discussion about the abilities and limits of model extrapolation in the context of describing climate change effects on species. The essence of the issue is that the models tend to make erratic or erroneous estimates when extrapolating outside the data limits on which they were trained. For example, if a model included January temperature as a variable, and presence/absence points included in the model had January temperatures ranging from -9.9—17.4 °C, the model could reasonably be expected to make predictions for sites with January temperature in that interval. Estimates of climate suitability may start to break down, though, when dealing with sites where January temperature is projected to reach temperatures greater than 17.4 °C because the model was not trained with these conditions, and thus it does not know how to classify sites with such extreme conditions.

It is easy to see this issue at play when using climate envelope models for climate change work, but the same thing can happen when using models with data from the same time period, but different areas spatially (for example, this may happen when using models to understand the potential distribution of invasive species). We may train a model in the native range of a species, say South America, and then use it to predict climate suitability for the species in the southern USA. It's easy to imagine that extrapolating the model trained in one area could result in 'new' values of climate data in the area for which you are making a 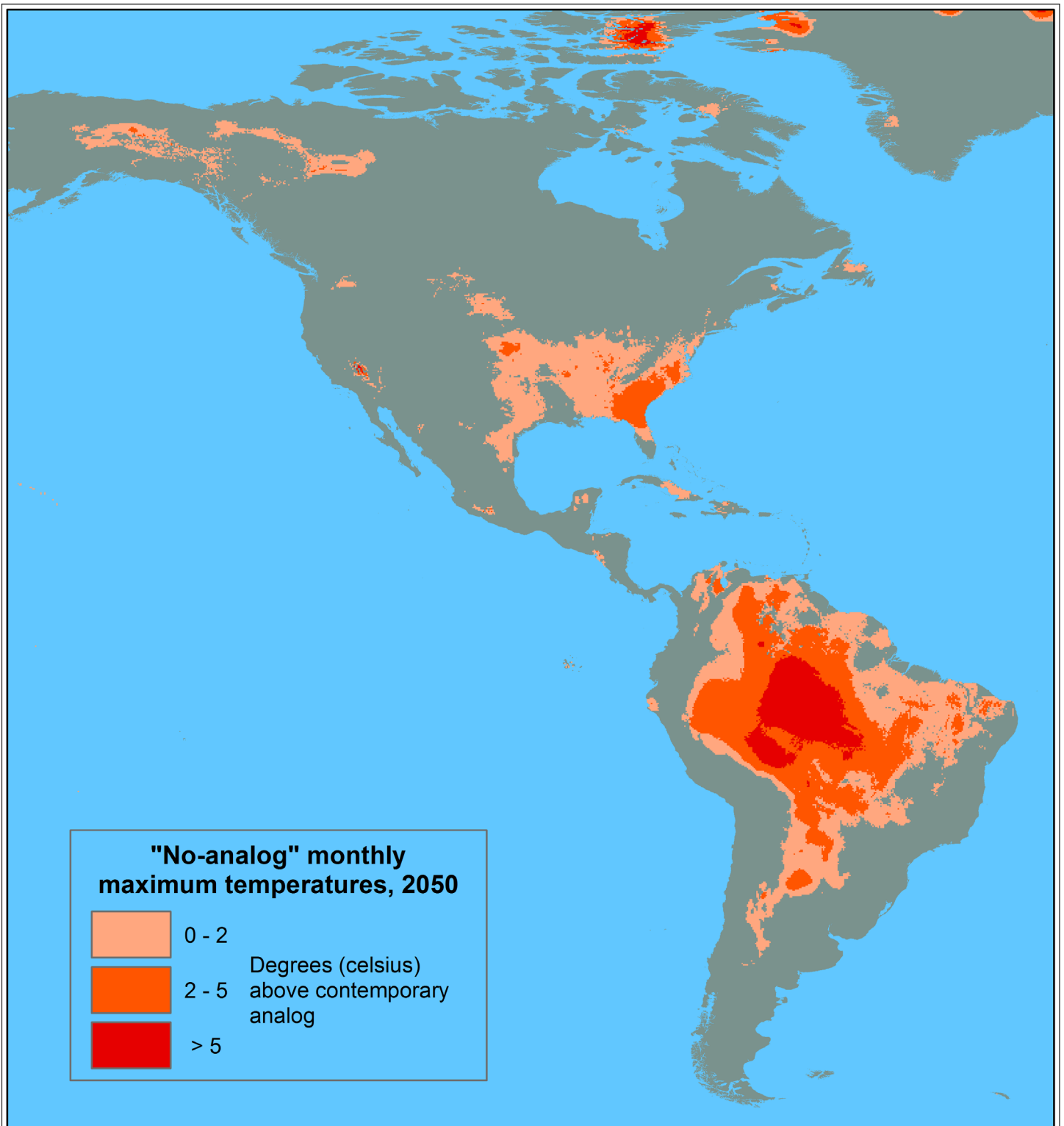map. As was the case when considering climate change effects, the implications for model predictions may be problematic.

There are several ways to deal with limiting the areas in which models make their predictions. The Maxent software for maximum entropy modeling uses a procedure called '**clamping**' to constrain model predictions to areas with climate (or other environmental conditions) within the range of values observed in the training data. More generally, an approach called MESS (**Multivariate Environmental Similarity Surface**) modeling can be used to investigate the appearance of novel climates in projection maps (Elith et al. 2010). It is definitely a good idea to take a conservative approach when doing climate envelope modeling, and restrict the area of model extrapolation to places where conditions are within the range of the calibration data.

What are 'no-analog' conditions?

When talking about climate change issues, it is not uncommon to hear people talking about '**no-analog**' conditions/climate/communities. The idea is conceptually related to the idea presented in the past section about extrapolating models into areas with conditions outside the boundaries for which the models were trained. As climate changes in a given location, new conditions may arise in that place that are different from what they are now, or were in the recent past (Figure 8). So an area that has long been characterized by two months of sub-freezing temperatures in the winter, for example, may be expected to only receive one or two hard freezes per year in the future. Thus, future conditions

"No-analog" monthly maximum temperatures, 2050

▮ (light)	0 - 2
▮ (orange)	2 - 5
▮ (red)	> 5

Degrees (celsius) above contemporary analog

This map provides an example of "no-analog" temperatures in the western hemisphere using a bioclimatic parameter (maximum temperature of warmest month). In 2050, areas in orange are expected to have hottest months with a mean daily maximum temperature greater than any experienced within a distance of 1000 kilometers under contemporary conditions (i.e., no contemporary analog for the future condition exists within 1000 km). This example uses climate projections from the Hadley Centre's HadCM3 GCM under the A2 IPCC emissions scenario.

Figure 8. Map illustrating the concept of no-analog climates.

are expected to be different from contemporary conditions such that no contemporary analog (or comparison) to the expected future condition exists. People talk about no-analog conditions at different scales, from regional to global. It may be possible, for example, that future climate conditions expected in one place may actually match climate at some other place today. If those two locations are on opposite sides of the Earth, however, people may still use the term 'no-analog' to refer to the future condition, because the matching climate condition is geographically very distant.

Among the implications of a no-analog future is that we can't expect natural systems in the future to look the way they do now, and they may not occur in the places they occur now. Because of this managing natural resources for future conditions may be quite a bit different than the management occurring today. In the context of climate envelope modeling, models may not perform particularly well when extrapolating into no-analog conditions, so users (and those of us interpreting results from climate envelope models) should remember to treat estimates in areas of no-analog climate with extra caution.

For more information:

Elith J., M. Kearney & S. Phillips. 2010. The art of modeling range-shifting species. Methods in Ecology and Evolution 1:330—342.

Williams J.W. & S. T. Jackson. 2007. Novel climates, no-analog communities and ecological surprises. Frontiers in Ecology and Environment 5:475—482.

Appendix I

Bioclimate variables
Annual Mean Temperature
Mean Diurnal Range (Mean of monthly (max temp - min temp))
Isothermality (annual mean temperature / mean diurnal range * 100)
Temperature Seasonality (variation across 12 months)
Max Temperature of Warmest Month
Min Temperature of Coldest Month
Temperature Annual Range (Max temperature of warmest month – min temperature of coldest month)
Mean Temperature of Wettest Quarter
Mean Temperature of Driest Quarter
Mean Temperature of Warmest Quarter
Mean Temperature of Coldest Quarter
Annual Precipitation
Precipitation of Wettest Month
Precipitation of Driest Month
Precipitation Seasonality (variation across 12 months)
Precipitation of Wettest Quarter
Precipitation of Driest Quarter
Precipitation of Warmest Quarter
Precipitation of Coldest Quarter

Appendix 1. List of 19 standard bioclimate variables.

Glossary

Algorithm: the mathematical function used to describe a species-climate relationship. Common algorithms in climate envelope modeling include generalized linear models, maximum entropy, and random forests.

AUC (Area under the Curve): The area under the receiver-operator characteristic curve is a threshold-independent metric of model performance. High values of AUC (closer to 1) indicate models that more accurately discriminate climate conditions at presence and absence sites.

Calibration: The process of fitting a species-climate relationship, usually based on contemporary climate data. Once a model is calibrated, its performance is evaluated and it can be extrapolated to new areas or time frames.

Clamping: A process in maximum entropy modeling wherein extrapolation of a model can be restricted to areas where climate conditions match those found in the calibration data set. Clamping ensures that models cannot be used to describe conditions outside the range for which they were calibrated (and for which predictions may be invalid).

Climate envelope model: A mathematical function describing climate suitability for a species based on a correlation between occurrence and climate. Climate envelope models are often extrapolated to describe future climate change effects on species. Climate envelope models are a subset of more general species distribution models.

Cohen's kappa (kappa): A threshold-dependent model evaluation metric that measures correctly and

incorrectly classified presences and absences which corrects for chance classifications. Cohen's kappa ranges from 0—1, with higher values indicating greater classification ability.

Collinearity: Intercorrelation between predictor variables in climate envelope models. Researchers often use a subset of relatively uncorrelated predictors, because high collinearity can alter predictions in climate envelope models.

Commission error: A classification error in which presence is falsely predicted. In a climate envelope model, commission error happens when a model predicts climate suitability for a species in an area where the species is not known to occur. The opposite of omission error.

Confusion matrix: A table in which the number of objects predicted to occur in a given number of categories is tallied with the number of objects actually occurring in the same categories. In a climate envelope model, a confusion matrix summarizes the number of species observations predicted and observed to occur in areas where climate is suitable or not.

Cross-validation: An approach to model evaluation often used when an independent dataset is not available to check the accuracy of a model. Alternative cross-validation approaches involve partitioning a large dataset into 'training' and 'testing' subsets. A model is created with the training data, and the ability of that model to classify the test data is evaluated using a cross-validation technique such k-fold partitioning or data subsetting.

Extrapolation: Application of a model outside the spatial or temporal boundaries within which it was created. Climate envelope models are often extrapolated to make projections about future climate change effects on species.

Generalized linear models: A widely-used algorithm for climate envelope modeling that builds from a linear regression framework to model variation in a dependent variable as a function of an independent predictor variable.

Map correlation: A metric of correspondence between prediction maps, in which each grid cell in a map is correlated with the corresponding grid cell in another map. Map correlation varies from 0—1, with higher values indicating greater correspondence between prediction maps.

Maximum entropy (Maxent): An algorithm for climate envelope modeling that calculates the probability of occurrence of a species as a function of environmental (climate) conditions where the species occurs relative to background conditions in the area of interest.

Multivariate Environmental Similarity Surface (MESS): An approach to identifying areas where climate conditions occur outside the boundaries of those for which a model was calibrated. Extrapolating results from climate envelope models in such areas is not recommended, and MESS provides a way to identify those areas that should be excluded when extrapolating models to project climate change effects on species.

Niche conservatism: A key assumption of climate envelope modeling that assumes species cannot adapt to new climate conditions different from those where it already occurs.

No-analog future: This phrase is used to describe the idea that climate change is likely to result in conditions in some places that are unlike those seen in that area in the past or present. Thus, there may be no present-day analog for expected future conditions that can be used to understand climate change responses.

Omission error: A classification error in which absence is falsely predicted. In a climate envelope model, omission error happens when a model fails to predict climate suitability for a species in an area where the species is known to occur. The opposite of commission error.

Overprediction: The tendency of a climate envelope model to predict climate suitability in areas where a species is not known to occur. Some overprediction is often inevitable, because species do not occur everywhere that climate is suitable. However, too much overprediction indicates a model is not effectively discriminating between suitable and unsuitable areas.

Prediction: An extrapolation of a climate envelope model into the near future for which validation data are likely to be available. Whereas a projection refers to an extrapolation that cannot be reasonably verified in a relatively short time frame, a prediction is an extrapolation that can be verified in the near future.

Projection: An extrapolation of a climate envelope model into a future time for which validation data are unavailable. Whereas a prediction refers to an extrapolation that can be reasonably verified in a relatively short time frame, a projection is an extrapolation that cannot be verified in the near future.

Pseudo-absences: Georeferenced point data used in many climate envelope modeling algorithms. Because

true biological absence is usually difficult to verify, researchers use 'pseudo-absences' where the presence of a species is unknown, along with the presence data where a species is known to occur.

Random forests: A modeling algorithm in which random subsets of selected climate predictors are used to classify presence and (pseudo-) absence over and over again in order to find values of climate predictors that best classify species occurrence.

Sensitivity: A model's ability to correctly classify species presences. In a model with high sensitivity, presences occur in grid cells where climate is categorized as suitable for a species.

Source-sink population dynamics: The idea that species may occur in areas (sinks) where the environment is unsuitable for long-term persistence because of subsidization from source populations. Species characterized by strong source-sink population dynamics violate a key assumption of climate envelope modeling, namely that species occur in areas suitable for long-term population persistence.

Species distribution model: A mathematical function describing environmental suitability for a species based on a correlation between occurrence and environmental conditions. Climate envelope models are a subset of more general species distribution models.

Specificity: A model's ability to correctly classify species absences. In a model with high specificity, absences occur in grid cells where climate is categorized as unsuitable for a species.

Threshold-dependent metrics: Approaches to model evaluation that require the user to define a threshold at which climate is categorized as suitable or

unsuitable. Cohen's kappa, sensitivity and specificity are all examples of threshold-dependent metrics. The raw data for calculating threshold-dependent metrics are often summarized in a confusion matrix.

Threshold-independent metrics: Approaches to model evaluation that do not require the user to define a threshold at which climate is categorized as suitable or unsuitable. The AUC is an example of threshold-independent metric.

Testing: In a cross-validation procedure, testing is done to evaluate a model created from the training dataset. Often, 25% of species occurrences (presences and absences) are used for model testing.

Training: In a cross-validation procedure, the training data are used to create a model, which is then evaluated using test data. Often, 75% of species occurrences (presences and absences) are used for model training.

Helpful Resources

http://www.worldclim.org/: Climate data online

http://www.prism.oregonstate.edu/: Climate data online

http://www.climatewizard.org/: Climate data online

http://www.ncdc.noaa.gov/oa/climate/normals/usnormals.html: Climate data online

http://www.ipcc.ch/: Intergovernmental Panel on Climate Change (IPCC) homepage

http://www.gbif.org/: Species occurrence database

http://zipcodezoo.com/: Species occurrence database

http://www.ornisnet.org/: Bird occurrence database

http://manisnet.org/: Mammal occurrence database

http://www.herpnet.org/: Amphibian and reptile occurrence database

https://www.pwrc.usgs.gov/BBS/: Breeding Bird Survey data

http://www.diva-gis.org/: Free GIS software

http://data.prbo.org/apps/bssc/: Climate change vulnerability for California birds

http://www.habitat.noaa.gov/pdf/scanning_the_conservation_horizon.pdf: Guide to climate change vulnerability assessment, including models and modeling

Literature Cited

Araújo M. B. & M. New. 2007. Ensemble forecasting of species distributions. Trends in Ecology & Evolution 22:42—47.

Araújo M. B. & A. T. Peterson. 2012. Uses and misuses of bioclimate envelope modeling. Ecology 93:1527—1539.

Austin M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modeling. Ecological Modelling 157:101—118.

Austin M. P. 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. Ecological Modelling 200:1—19.

Bidinger K., S. Lotters, D. Rodder & M. Veith. 2012. Species distribution models for the alien invasive Asian Harlequin ladybird (*Harmonia axyridis*). Journal of Applied Entomology 136:109—123.

Bray D. & H. von Storch. 2009. "Prediction" or "projection"?: the nomenclature of climate science. Science Communication 30:534—543.

Buckley L.B., M. C. Urban, M. J. Angilletta, L. G. Crozier, L. J. Rissler & M. W. Sears. 2010. Can mechanism inform species distribution models? Ecology Letters 13:1041—1054.

Chefaoui R.M. & J. M. Lobo. 2008. Assessing the effects of pseudo-absences on predictive distribution model performance. Ecological Modelling 210:478—486.

Cutler D. R., T. C. Edwards Jr., K. H. Beard, A. Cutler, K. T. Hess, J. Gibson & J. J. Lawler. 2007. Random forests for classification in ecology. Ecology 88:2783—2792.

Dormann C. F., O. Purschke, J. R. García Márquez, S. Lautenbach & B. Schröder. 2008. Components of uncertainty in species distribution analysis: a case study of the Great Grey Shrike. Ecology 89:3371—3386.

Elith J., M. A. Burgman & H. M. Regan. 2002. Mapping epistemic uncertainties and vague concepts in predictions of species distribution. Ecological Modelling 157:313—329.

Elith J. & C. H. Graham. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography 32:66—77.

Elith J., C. H. Graham, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. McC. Overton, A. Townsend Peterson, S. J. Phillips, K. Richardson, R. Schachetti-Pereira, R. E. Schapire, J. Soberón, S. Williams, M. S. Wisz, and N. E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129—151.

Elith J., M. Kearney & S. Phillips. 2010. The art of modeling range-shifting species. Methods in Ecology and Evolution 1:330—342.

Elith J., S. J. Phillips, T. Hastie, M. Dudík, Y. E. Chee & C. J. Yates. 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17:43—57.

Fielding A. H. & J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24:38—49.

Franklin J. 2009. Mapping species distributions: spatial inference and prediction. Cambridge University Press, New York.

Freeman E. A. & G. G. Moisen. 2008. A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. Ecological Modeling 217:48—58.

Glick P., B. A. Stein & N. A. Edelson. 2011. Scanning the conservation horizon: a guide to climate change vulnerability assessment. National Wildlife Federation, Washington DC.

Guisan A., N. E. Zimmermann, J. Elith, C. H. Graham, S. Phillips & A. T. Peterson. 2007. What matters most for predicting the occurrence of trees: techniques, data, or species' characteristics? Ecological Monographs 77:615—630.

Hernandez P. A., C. H. Graham, L. L. Master & D. L. Albert. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29:773—785.

Hijmans, R. S., S. E. Cameron, J. L. Parra, P. G. Jones & A. Jarvis. 2005. Very high resolution climate surfaces for global land areas. International Journal of Climatology 25:1965—1978.

Kearney M. P. & W. Porter. 2009. Mechanistic niche modeling: combining physiological and spatial data to predict species ranges. Ecology Letters 12:334—350.

Kearney M. P., B. A. Wintle & W. P. Porter. 2010. Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. Conservation Letters 3:203—213.

LaSorte F. A. & W. Jetz. 2012. Tracking of climate niche boundaries under recent climate change. Journal of Animal Ecology 81:914—925.

Lawler J.J., S. L. Shafer, D. White, P. Kareiva, E. P. Maurer, A. R. Blaustein & P. J. Bartlein. 2009. Projected climate-induced faunal change in the western hemisphere. Ecology 90:588—597.

Lobo J. M., A. Jímenez-Valverde & J. Hortal. 2010. The uncertain nature of absences and their importance in species distribution modeling. Ecography 33:103—114.

Lobo J. M., A. Jímenez-Valverde & R. Real. 2007. AUC- a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography 17:145—151.

Manel S., H. C. Williams & S. J. Omerod. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. Journal of Applied Ecology 38:921—931.

McCullugh P. & J. A. Nelder. 1989. Generalized Linear Models. Chapman and Hall, London.

Meynard C. N. & J. F. Quinn. 2007. Predicting species distributions: a critical comparison of the most common statistical models using artificial species. Journal of Biogeography 34:1455—1469.

New M., D. Lister, M. Hulme & I. Makin. 2002. A high-resolution data set of surface climate over global land areas. Climate Research 21:1—25.

Nix H. A. 1986. A biogeographic analysis of Australian elapid snakes. In: Longmore R., Atlas of elapid snakes of Australia. Australia flora and fauna series 7. Bureau of Flora and Fauna, Canberra.

Pagel J. & F. M. Schurr. 2012. Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. Global Ecology and Biogeography 21:293—304.

Parra J. L. & W. B. Monahan. 2008. Variability in 20th century climate change reconstructions and its consequences for predicting geographic responses of California mammals. Global Change Biology 14:2215—2231.

Phillips S. J. & M. Dudík. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31:161—175.

Phillips S. J., R. P. Anderson & R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190:231—259.

Phillips S. J., M. Dudík, J. Elith, C. H. Graham, A. Lehmann, J. Leathwick & S. Ferrier. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological Applications 19:181—197.

Platts P. J., A. Ahrends, R. E. Gereau, C. J. McClean, J. C. Lovett, A. R. Marshall, P. K. E. Pellikka, M. Mulligan, E. Fanning & R. Marchant. 2010. Can distribution models help refine inventory-based estimates of conservation priority? A case study in the Eastern Arc forests of Tanzania and Kenya. Diversity and Distributions 16:628—642.

Pulliam H. R. 1988. Sources, sinks, and population regulation. American Naturalist 132:652—661.

Schurr F. M., J. Pagel, J. Sarmento Cabral, J. Groeneveld, O. Bykova, R. B. O'Hara, F. Hartig, W. D. Kissling, H. P. Linder, G. F. Midgley, B. Schröder, A. Singer & N. E. Zimmermann. In Press. How to understand species' niches and range dynamics: a demographic research agenda for biogeography. Journal of Biogeography.

Souther S. & J. B. McGraw. 2011. Evidence of local adaptation in the demographic response of American ginseng to interannual temperature variation. Conservation Biology 25:922—931.

Syphard A. D. & J. Franklin. 2009. Differences in spatial predictions among species distribution modeling methods vary with species traits and environmental predictors. Ecography 32:907—918.

VanDerWal J., L. P. Shoo, C. Graham & S. E. Williams. 2009. Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? Ecological Modelling 220:589—594.

Vitasse Y, S. Delzon, C. C. Bresson, R. Michalet &

A. Kremer. 2009. Altitudinal differentiation in growth and phenology among populations of temperate-zone tree species growing in a common garden. Canadian Journal of Forest Research 39:1259—1269.

Watling J. I., S. S. Romañach, D. N. Bucklin, C. Speroterra, L. A. Brandt, L. G. Pearlstine & F. J. Mazzotti. 2012. Do bioclimate variables improve performance of climate envelope models? Ecological Modelling 246:79—85.

Weber S. B., A. C. Broderick, T. G. G. Groothuis, J. Ellick, B. J. Godley & J. D. Blout. 2012. Fine-scale thermal adaptation in a green turtle nesting population. Proceedings of the Royal Society B-Biological Sciences 279:1077—1084.

Wiens J. A., D. Stralberg, D. Jongsomjit, C. A. Howell & M. A. Snyder. 2009. Niches, models, and climate change: assessing the assumptions and uncertainties. Proceedings of the National Academy of Sciences USA 106:19729—19736.

Williams J.W. & S. T. Jackson. 2007. Novel climates, no-analog communities and ecological surprises. Frontiers in Ecology and Environment 5:475—482.

Wisz M. S., R. J. Hijmans, J. Li, A. T. Peterson, C. H. Graham, A. Guisan & NCEAS Predicting Species Distributions Working Group. 2008. Effects of sample size on the performance of species distribution models. Diversity and Distributions 14:763—773.