Author/s:
Merow, C;Smith, MJ;Edwards, TC;Guisan, A;Mcmahon, SM;Normand, S;Thuiller, W;Wüest, RO;Zimmermann, NE;Elith, J

Title:
What do we gain from simplicity versus complexity in species distribution models?

Date:
2014-12-01

Citation:
Merow, C., Smith, M. J., Edwards, T. C., Guisan, A., Mcmahon, S. M., Normand, S., Thuiller, W., Wüest, R. O., Zimmermann, N. E. & Elith, J. (2014). What do we gain from simplicity versus complexity in species distribution models?. Ecography, 37 (12), pp.1267-1281. https://doi.org/10.1111/ecog.00845.

Persistent Link:
https://hdl.handle.net/11343/52663

Back to the basics of species distribution modeling: what do we gain from complex versus simple models?

Cory Merow[1,2,3,*], Mathew J. Smith[3], Thomas C. Edwards, Jr.[4], Antoine Guisan[5,6], Sean M. McMahon[1], Signe Normand[7], Wilfried Thuiller[8,9], Rafael O. Wüest[7,8,9], Niklaus E. Zimmermann[7], Jane Elith[10]

1 Smithsonian Environmental Research Center, Edgewater, MD 21307-0028, USA

2 University of Connecticut. Ecology and Evolutionary Biology. 75 North Eagleville Rd. Storrs, CT, 06269, USA

3 Computational Science Laboratory, Microsoft Research, CB1 2FB, UK

4 U.S. Geological Survey, Utah Cooperative Fish and Wildlife Research Unit and Department of Wildland Resources, Utah State University, Logan, UT 84322 USA

5 Departement d'Ecologie et d'Evolution, Université de Lausanne, CH-1015 Lausanne, Switzerland

6 Institute of Earth Surface Dynamics, Université de Lausanne, CH-1015 Lausanne, Switzerland

7 Landscape Dynamics, Swiss Federal Research Institute WSL, Zürcherstr. 111, CH-8903 Birmensdorf, Switzerland

8 Univ. Grenoble Alpes, LECA, F-38000 Grenoble, France

9 CNRS, LECA, F-38000 Grenoble, France

[10] Centre of Excellence for Biosecurity Risk Analysis; School of Botany, The University of Melbourne, Parkville 3010 Australia

* Corresponding author email: cory.merow@gmail.com

Keywords: niche modeling, nonparametric models, parametric models, species range, biogeography, machine learning.

ABSTRACT

Species distribution models (SDMs) are widely used to explain and predict species ranges and environmental niches. They are most commonly constructed by inferring species' occurrence-environment relationships using statistical and machine-learning methods. The variety of methods that can be used to construct SDMs (e.g., generalized linear/additive models, tree-based models, maximum entropy, etc.), and the variety of ways that such models can be implemented, permits substantial flexibility in SDM complexity. Building models with an appropriate amount of complexity for the study objectives is critical for robust inference. By characterizing complexity as the shape of the inferred occurrence-environment relationships and the number of parameters used to describe them, we can gain insights into whether additional complexity is valuable or superfluous. By building 'under fit' models, having insufficient flexibility to describe observed occurrence-environment relationships, we risk misunderstanding the factors shaping species distributions. By building 'over fit' models, with excessive flexibility, we risk inadvertently ascribing pattern to noise or building opaque models. However, model selection can be challenging, especially when comparing models constructed under different modeling paradigms. Here we argue for a more pragmatic approach: researchers should constrain the complexity of their models based on study objective, attributes of the data, and an understanding of how these interact with the underlying biological processes. We discuss guidelines for balancing under fitting with over fitting and consequently how complexity affects decisions made during model building. Although some generalities are possible, our discussion reflects differences in opinions that favor simpler versus more complex models. We conclude that combining insights from both simple and complex SDM building approaches best advances our knowledge of current and future species ranges.

I. INTRODUCTION

Species Distribution Models (SDMs), also known as ecological niche models or habitat selection models, are widely used in ecology, evolutionary biology, and conservation (Elith and Leathwick 2009, Franklin 2010, Zimmermann et al. 2010, Peterson et al. 2011, Svenning et al. 2011, Guisan et al. 2013). SDMs can provide insights into generalities and idiosyncrasies of the drivers of complex patterns of species' geographic distributions. SDMs are built using a variety of statistical methods – e.g., generalized linear/additive models, tree-based models, maximum entropy – which span a range of complexity in the occurrence-environment relationships that they fit. Capturing the appropriate amount of complexity for particular study objectives is challenging. By building 'under fit' models, having insufficient flexibility to describe observed occurrence-environment relationships, we risk misunderstanding the factors shaping species distributions. By building 'over fit' models, with excessive flexibility, we risk inadvertently ascribing pattern to noise or building opaque models. As such, determining a suitable amount of complexity to include in SDMs is crucial for biological applications. Because traditional model selection is challenging when comparing models from different SDM modeling paradigms (e.g. those in Table 1), we argue that researchers must constrain model complexity based on attributes of the data and study objectives and an understanding of how these interact with the underlying biological processes. Here, we discuss the challenges that choosing an appropriate among of model complexity poses and how this influences the use of different statistical methods and modeling decisions (Elith and Graham 2009).

Complexity is a fundamental part of observed occurrence patterns because occurrence-environment relationships may be confounded by processes that are not explicitly related to the environment, such as competition, dispersal, response to disturbance, and biotic interactions (Pulliam 2000, Holt 2009, Boulangeat et al. 2012). Describing this complexity is

71 critical for many applications of SDMs, and using flexible occurrence-environment
72 relationships allows biologists to hypothesize about the drivers of complexity or make
73 accurate predictions that derive from their representation in SDMs. However, building
74 complex models comes with the challenge of differentiating true complexity from noise (see
75 Chapter 2 in (Hastie et al. 2009)). Skepticism persists regarding whether flexible models are
76 often over-fit to noise that is prevalent in many occurrence data sets. This skepticism has
77 fueled interest in developing the array of statistical methods and modeling decisions that,
78 taken together represent the different modeling paradigms that we discuss below (Table 1).
79 We characterize model complexity by the shape of the inferred occurrence-environment
80 relationships (Table 1) and the number of posited predictors and parameters used to describe
81 them. We do not focus on quantitative measures of complexity, such as the effective degrees
82 of freedom, that might apply to the spectrum of modeling paradigms considered here, as
83 these have been shown to be unreliable in characterizing model complexity (Janson et al.
84 2013). Rather, we turn to tools that visualize potentially high dimensional occurrence-
85 environment relationships via univariate 'response curves.' The most common approach is to
86 plot the predicted occurrence probability against the predictor of interest by holding all other
87 predictors at their mean or median values (Elith et al. 2005; see Table 1), although other
88 approaches are possible (Fox 2003, Hastie et al. 2009). Though insightful, this imperfectly
89 represents the true fitted because one cannot visualize the effects of interactions in such plots
90 (3-dimensional response surfaces or the "inflated response curves" of Zurell et al (2012) help
91 here). Such responses must be interpreted as conditional on the other mean or median
92 predictors in the model, which may be different than the responses to variables held at other
93 values (Zurell et al. 2012), or to an unconditional model. Nevertheless, when interpreted
94 appropriately, response curves can provide valuable insights into the drivers of occurrence
95 patterns.

In this paper, we develop general guidelines for deciding on an appropriate level of complexity in occurrence-environment relationships. Uncertainty about how best to describe ecological complexity has to some extent divided biologists between those who prefer to use the principle of parsimony to identify model complexity (preferring the simplest model that is consistent with the data), and those that try to approximate more of the complexities of the real world relationships. We review the literature and general modeling principles emerging from these two viewpoints, and discuss the ways in which these overlap or differ in light of study objectives and attributes of the data. We make a variety of recommendations for choosing levels of complexity under different circumstances, while highlighting unresolved scenarios where viewpoints differ. We conclude with suggestions for drawing from the strengths of each modeling paradigm in order to advance our knowledge of current and future species geographical ranges.

A. COMPLEXITY IN ECOLOGY

Many interacting biotic and abiotic processes influence species distributions and can manifest as complex occurrence-environment relationships (Soberón 2007, Boulangeat et al. 2012). The essential challenge to recovering primary environmental drivers of these distributions, however, is to differentiate the signals of range determinants from sampling and environmental noise. Before embarking on statistical analyses of range determinants, ecological theory can focus an investigation (Austin 1976, Pulliam 2000, Austin 2002, Chase and Leibold 2003, Austin 2007, Holt 2009). There are, *a priori,* a set of common drivers of populations that can be used to propose general shapes of occurrence-environment relationships. For example, we expect that for many variables, response curves describing a fundamental niche should be smooth because sudden jumps in fitness along an environmental gradient are unlikely to exist (Pulliam 2000, Chase and Leibold 2003, Holt 2009). For other variables, e.g. related to thermal tolerance, steep thresholds may exist due to loss of

121 physiological function (Buckley et al. 2011). However, response curves describing realized
122 niches might exhibit discontinuities due to the multiple interacting factors that can limit a
123 species' occurrence in any particular location. Unimodal responses are expected (e.g. a bell-
124 shaped curve) because conditions too extreme for survival often exist at either end of a
125 proximal gradient (Austin 2007). However, response curves can be linear where only part of
126 the environmental range of the species has been sampled (e.g. one side of a unimodal
127 response; Albert et al. 2010). Austin and Smith's (1989) continuum concept for plant species
128 distributions predicts that skewed unimodal response curves are likely when plant species
129 distributions are predominantly determined by one or a few environmental variables that
130 strongly regulate survivorship and or reproduction (e.g. by temperature thresholds), but that
131 more irregular response curves are expected given that species are influenced by a range of
132 regulatory factors (e.g. different limiting nutrients, biotic and abiotic interactions) and
133 historical contingencies (Austin et al. 1994, Normand et al. 2009). Even with single factors,
134 the processes that determine fitness may be different across the range, e.g., where one
135 temperature extreme leads to abrupt loss of function while the other extreme causes gradually
136 reduced performance. Interaction terms can be desirable to capture covariation between
137 predictors or tradeoffs along resource gradients (e.g. higher temperatures are tolerable with
138 greater rainfall). Many applications of SDMs do not explicitly consider such theoretical
139 constraints on the shape of response curves (but see Santika and Hutchinson 2009).
140 Therefore, we are faced, fundamentally, with the challenge of inferring unknown levels of
141 ecological complexity through the lens of models that imperfectly capture it.

142 **B. COMPLEXITY IN MODELS**

143 Two attributes of model fitting determine the complexity of inferred occurrence-
144 environment relationships in SDMs: the underlying statistical method and modeling decisions

made about inputs and settings. Together these define different modeling paradigms, a number of which are illustrated in Table 1.

1. Statistical Methods

One of the primary differences among the available statistical methods for fitting SDMs is the range of transformations of predictors that they typically consider (in machine learning parlance: which "features" to allow), and this helps to define the upper limit of complexity for their fitted response surfaces. We detail commonly used modeling paradigms and demonstrate examples of their response curves in Table 1. Rectilinear or convex-hull environmental envelopes (e.g., BIOCLIM or DOMAIN) and distance-based approaches in multivariate environmental spaces (e.g. Malahanobis) are used in the simplest SDMs. Their response curves are simple functions (e.g., linear, hinge or step; Elith et al. 2005).Generalized linear models (GLMs), which are typically fitted with linear or polynomial features up to second order terms (rarely third or fourth order) for SDMs, and often without interactions, admit more complexity. Generalized additive models (GAMs) are potentially more complex because they allow non-parametric smooth functions of variable flexibility (Hastie and Tibshirani 1990, Wood 2006). Decision trees (Breiman et al. 1984) can also become quite complex because these can use a large number of step functions (each requiring a parameter) and can implicitly include high order interaction terms to depict response curves of arbitrary complexity.

2. Modeling Decisions

Decisions that affect model complexity apply to all the statistical methods described above. For example, if a large set of predictors are available, then model complexity will differ depending on whether the full set, or a small subset, is used. One must also determine which features are considered in the model. Each feature requires at least one parameter in the occurrence-environment relationship and hence increases model complexity. Large

numbers of predictors are more commonly used in machine-learning approaches because they automate feature selection whereas fewer are often used in simpler models where features are specified *a priori*. For example, maximum entropy models (MAXENT) can consider any number of linear, quadratic, product, threshold (step functions) or hinge transformations of the predictors (Phillips et al. 2006). In principle, this same complexity could be fit in a traditional GLM but this is typically impractical and not of interest to ecologists.

SDM complexity is amplified when interactions between predictors are included. Interactions between predictors are valuable because they account for the high correlation between many environmental drivers. GLMs and GAMs can include interactions that have been specified during model formulation as potentially ecologically relevant, but are usually used only sparingly. Decision trees include interactions implicitly through their hierarchical structure; i.e., the response to one variable depends on values of inputs higher in the tree, meaning that high order interaction terms (that depend on all the predictors along a branch) are possible. However interactions between variables are selected automatically if supported by the data.

Using ensembles of models can increase or decrease complexity. Ensembles are combinations of models in which the component models can be chosen based on selected criteria (e.g. predictive performance on held out data; Araújo and New 2007) or with an ensemble algorithm (a machine learning method). For instance, regression models selected via an information criterion can be combined using "multi-model inference", allowing distributions over effect sizes and over predictions to new sites (Burnham and Anderson 2002). A typical machine learning approach to ensembles uses an algorithm to build an ensemble of simple models that together predict better than any one component model. Examples include bagging and boosting - these can be used on any component models; in ecology the most used component models are decision trees (e.g., in Random Forests,

195 Brieman 2001; and Boosted Regression Trees, Friedman 2001). Bagging (bootstrap
196 aggregation) can be used to fit many models to bootstrapped replicates of the dataset (with
197 and without random subsetting of predictors used across trees as in Random Forests). In
198 contrast, boosting uses a forward stage wise method to build an ensemble, at each step
199 modeling the residuals of the models fitted to date. Taking ensembles of relatively simple
200 models can increase complexity because combinations of simple models will not necessarily
201 be simple. In contrast, ensembles of more complex models can average over idiosyncrasies of
202 individual models to produce smoother response curves (Elder 2003).

203 *3. Model comparison*

204 Comparing models across modeling paradigms (e.g. those in Table 1) can be challenging.
205 This is one of our motivations for constraining model complexity based on study objectives
206 and data attributes. Information theoretic measures are a conventional way to choose model
207 complexity and are relatively easy to apply for models where estimating the number of
208 degrees of freedom is possible. However these cannot be calculated for ensemble-based
209 methods nor for many other methods in common use (Janson et al. 2013). In fact, Janson et
210 al. (2013) warn, "contrary to folk intuition, model complexity and degrees of freedom are *not*
211 synonymous and may correspond very poorly". One way to compare models produced by
212 different algorithms is to adopt a common currency for model performance by evaluating
213 model predictions on either the training data or independent testing data. Measures such as
214 AUC, Cohen's Kappa, and the True Skill Statistic are based on correctly distinguishing
215 presences from absences. Measures based on non-thresholded predictions are also relevant
216 and favorable in many situations (Lawson et al. 2013). However, each of these metrics has
217 weaknesses in different circumstances (Lobo et al. 2008) and further, only represent heuristic
218 diagnostics for presence-only data, because presences must be compared to
219 pseudoabsence/background data (Hirzel et al. 2006).

220 Once one has determined a suitable modeling paradigm, tuning of the amount of
221 complexity is more straightforward using a range of model selection techniques. Feature
222 significance (e.g., p-values), measures of model fit (e.g., likelihood), and information criteria
223 (e.g., AIC, AICc, BIC; Burnham and Anderson 2002) can be applied to regression-based
224 methods. Cross-validation can be or other resampling techniques are also used to set the
225 smoothness of splines in GAMs (Wood 2006) or determine tuning parameters in most
226 machine learning methods (Hastie et al. 2009). Shrinkage or regularization is used in
227 regression, MAXENT and boosted regression trees to constrain coefficient estimates so
228 models predict reliably (Phillips et al. 2006, Hastie et al. 2009). Loss functions, which
229 penalize for errors in prediction, can be constructed for any of the modeling paradigms we
230 consider (Hastie et al. 2009). An alternative approach employs null models to evaluate
231 whether additional complexity has lead to spurious predictive accuracy (Raes and ter Steege
232 2007).
233 Evaluation against fit to training data alone cannot control for over fitting and risks
234 selecting excessively complex models (Pearce and Ferrier 2000, Araújo et al. 2005). In
235 general, best practice involves splitting the data into training data to fit the model, validation
236 data for model selection, and test data to evaluate the predictive performance of the selected
237 model (Hastie et al. 2009). Recent studies have emphasized that care should also be taken in
238 how data is partitioned into training, evaluation and test data, in particular to control for
239 spatial autocorrelation (Dormann et al. 2007; Latimer et al. 2007; Veloz, 2009; Hijmans
240 2012; see section II.C.5 for more details). Hence methods such as block cross-validation
241 (where blocks are spatially stratified) are gaining momentum (e.g., Hutchinson et al. 2011,
242 Pearson et al. 2013, Warton et al. 2013). Failure to factor out spatial autocorrelation in data
243 partitioning can lead to misleadingly good estimates of model predictive performance.

Basing model comparison on holdout data presents some practical challenges. Sample size may be insufficient to subset the data without introducing bias. Subsets of data can contain the same or different biases compared to the full data set. In particular, it can be difficult to remove spatial correlation between training and holdout data when the sampling design for the occurrence data is unknown or when a species is restricted geographically or environmentally (this is discussed in section II.C.2 below). Removing autocorrelation can also lead to a drastic reduction in sample size (Veloz 2009, Hijmans 2012).

Importantly, all these approaches to model comparison have strengths and weaknesses and none can unambiguously select between models of differing complexity built with different statistical methods and underlying assumptions. The tried and tested methods of statistics and machine learning for model selection are valuable when working within a particular modeling paradigm, but to benefit from these, it is beneficial to narrow the scope of the feasible models based on biological considerations. A motivation for this paper is to discuss approaches for identifying appropriate level of complexity for particular study objectives based on data limitations and the underlying biological processes.

II. Philosophical, statistical and biological considerations when choosing complexity

In this section, we discuss the factors that can influence the choice of model complexity. First, we outline general considerations and philosophical differences underlying both simple and complex modeling strategies (section II.A). Next, we discuss how the study goals (section II.B) and data attributes (section II.C) and interact with model complexity. Figure 1 summarizes our findings. Importantly, a general consensus for choosing model complexity is lacking in many cases. To reflect the different schools of thought, we divide our recommendations in to that are relatively uncontroversial (subsections denoted

267 '**Recommendations**'), favor simple models (denoted '**Simple**'), and favor more complex
268 models (denoted '**Complex**')

269 **A. Simple versus complex: Fundamental approaches to describing natural systems**

270 ***Simple:*** Simple models tend towards a conservative, parsimonious approach and typically
271 avoid over-fitting. They link model structure to hypotheses that posit occurrence-environment
272 relationships *a priori* and examine whether the resulting model meets these expectations.
273 Simple models have greater tractability, can facilitate the interpretation of predictors (cf.
274 Tibshirani 1996), can help in understanding the primary drivers of species occurrence
275 patterns, and are likely to be more easily generalized to new data sets (Randin et al. 2006,
276 Elith et al. 2010). Although complex responses surely exist in nature, we cannot often detect
277 them because their signal is weak or they are confounded with sampling noise, bias or
278 autocorrelation. By using models that are too complex, one can inadvertently assign patterns
279 due to either data limitations or missing processes, or both, to environmental suitability and
280 fit the patterns simply by chance.

281 ***Complex:*** Complex models are often semi- or fully non-parametric, and can be preferred
282 when there is no desire to impose parametric assumptions, specific functional forms or pre-
283 select predictors for models *a priori*. This does not mean that they are not biologically
284 motivated, but more emphasizes the reality that nature as we observe it is a complex
285 phenomenon. Simple models may be readily interpretable but misleading (Breiman 2001),
286 and for many applications of SDMs a preference for predictive accuracy in new data sets over
287 interpretability is justifiable. Also, complex models are not necessarily difficult to interpret.
288 Indeed, their complexity can be valuable for suggesting novel, unexpected responses. If we
289 do not explore the full spectrum of complexity, there is a risk of obtaining an overly
290 simplified, or even biased, view of ecological responses. Complex models can, depending on
291 how they are structured, still identify simple relationships if responses are strong and robust.

B. Study objectives

1. Niche Description vs. Range Mapping

Two prominent applications of SDMs are characterizing the predictors that define a species' niche and projecting fitted models across a landscape. Niche characterization quantifies the variables, primarily climate and physical, that affects a species' distribution. This is often done by analyzing response curves, the functions (coefficients or smoothing terms) that define them, and their relative importance in the model. Projecting these fitted models across a landscape can predict the geographic locations where the species may occur in the present or in the future. In some studies, focus lies in the final mapped predictions rather than how they derive from the underlying fitted models.

Recommendations: Some evaluation of the biological plausibility of the shape and complexity of response curves is always valuable, even if the objective is not niche description. Such evaluation is particularly critical for extrapolation (section II.B.3), though it is admittedly quite challenging in multivariate models. Modelers should also carefully evaluate whether maps built from complex models substantially differ from maps built from simple models. If the predictions differ, the source of this should be explored. If the interest lies in interpretation, it is important to assess whether the mapped predictions are right for the right reason, and that complex environmental responses have not become proxies for sources of spatial aggregation in the data that lead to bias when projected to other locations (whether interpolation or extrapolation; section II.C.5).

Simple: Simple models are preferable for niche description because they usually yield straightforward, smooth response curves that can be linked directly to ecological niche theory (section I.B; Austin 2007), in contrast to the often irregular shapes that result from complex models (Table 1). Assumptions about species responses are more transparent when simple models are being projected in new situations.

***Complex**:* Complex models can be valuable for describing a species' niche when only qualitative descriptors of response curves are necessary (e.g. positive/negative, modality, relative importance) – i.e. even complex responses can be described in terms of main trends. Allowing complexity might offer more chance of identifying relevant response shapes. Complex models can be powerful for accurately mapping within the fitting region (Randin et al. 2006, Elith et al. 2006) when one is not necessarily concerned with an ecological understanding of the complexity of underlying models. Although the source of complex relationships may remain unknown, complex models have the flexibility to describe these. Abrupt steps in response curves might be helpful to uncover strictly unsuitable sites when mapping distribution in space.

2. Hypothesis Testing vs. Hypothesis Generation

Some SDM studies are focused on testing specific hypotheses related to how species are distributed in relation to particular predictors or features. In others, little is known about the predictors shaping the distribution and the objective is to explore occurrence-environment relationships and generate hypotheses for explanation. The approach critically affects the level of complexity permitted because hypothesis testing depends on being able to isolate the affects of particular features, whereas this matters less when exploring data in order to generate hypotheses.

Recommendations: When testing hypotheses, insights from ecological theory can guide the selection of features to include. A higher degree of control over the specific details of the underlying response surface is likely to be needed for hypothesis testing, which is made much easier using simple models. Hypothesis testing is more challenging in complex models with correlated features that can trade off with one another. Complex models are well suited to hypothesis generation, enabling a wider range of environmental covariates and modeling options than can be conveniently explored with simple models.

342 ***Simple:*** When the goal is hypothesis testing, simple parametric models allow
343 investigation of the strength and shape of relationships between species occurrence and a
344 small set of features. Furthermore, parametric models allow for hypothesis tests to examine if
345 specific nonlinear features should be included in the selected model(s). The problem with
346 complex models in this setting is that with the large suite of potential features that they use, it
347 is challenging to determine the significance of a single feature or attribute of the response
348 curve or to compare alternative models. Instead, one is constrained to accept the features
349 selected by the statistical method (e.g. features classes in MAXENT; splits in tree-based
350 methods) to represent that predictor (within some user-specified bounds). Rather, it is
351 preferable to specify a set of features (or multiple sets for competing models) to determine
352 the suitability for describing a particular pattern. For example, when features are selected
353 automatically, it may be challenging to determine whether a quadratic term that makes the
354 response unimodal is important or how much better/worse the model might be without it.
355 ***Complex****:* The starting premises, for hypothesis testing, is *a priori* ecological
356 understanding enabling the user to select a small set of features. However, we do not always
357 have this prior understanding. Complex models explore much larger sets of nonlinear features
358 and interactions than simple models and are suited for generating hypotheses about
359 underlying processes (Boulangeat et al. 2012) derived from potentially flexible responses that
360 would not often be detected with simpler models (e.g. bimodality). This same flexibility can
361 be used to augment existing knowledge. For example, if we know that a species is associated
362 with dry, high elevation locations, we don't need a simplified model to describe this, but
363 rather more insight from a potentially complex model to capture bimodality or strong
364 asymmetries. Complex models also provide tools for evaluating predictor importance, which
365 is useful for testing hypotheses and can lead to inference that differs little from simpler
366 models (Grömping 2009). These importance indices can be generated from permutation tests

(Strobl et al. 2008, Grömping 2009), contribution to the likelihood (e.g. 'percent contribution' in MAXENT), or proportion of deviance explained (decision trees).

3. Interpolate vs. Extrapolate

When predicting species' distributions over space and time, it is important to distinguish between interpolation and extrapolation. When a point is interpolated by a fitted model, it lies within the known data range of predictors, but was not measured for its response. Alternatively, an extrapolated point is one that lies outside the observed range of the predictor. Both interpolation and extrapolation can occur in geographic space or environmental space (cf. Peterson et al. 2011, Aarts et al. 2012). Process-based models are generally better for extrapolation because the functional form of the response curve captures the processes that apply beyond the range of observed data (Kearney and Porter 2009; Thuiller et al. 2013). Extrapolation requires caution in all scenarios but cannot be avoided when assessing questions relating to 'no-analogue' climate scenarios (Araújo et al. 2005) or range expansion.

Recommendations: The challenges associated with interpolation and extrapolation, though differing in the way they manifest, are apparent for models of any complexity and hence simple and complex perspectives align. Interpolation within the range of the observed data will be accurate if the model includes all processes operating in the interpolation extent and is based on well-structured data. Without that, prediction to unsampled sites will average across unrepresented processes and might reflect biases in the sample. More generally, it may not matter whether a response curve is complex as long as it retains the basic qualities of a simpler model. For example, a line or a sequence of small step functions parallel to the line can produce similar predictions. Some caution should be taken with complex models, as complex combinations of features can become proxies for unmeasured spatial factors in

391 unintended ways and inadvertently model clustering in geographic space as complexity in
392 environmental space, which can lead to errant interpolation (section II.C.5).
393 Extrapolation always requires that response curves have been checked for biological
394 plausibility (cf. section II.B.1). Of course, even simple models can extrapolate poorly. For
395 example, Thuiller et al. (2004) showed that a simple GLM or GAM run on a restricted and
396 incomplete range could create spurious termination of the smoothed relationships, leading to
397 errant extrapolation. Hence, the importance of extrapolation can depend on the chosen spatial
398 extent and on the selected features (section II.C.4). Complex models should be carefully
399 monitored at the edges of the data ranges, both due to small sample sizes there and the ways
400 different statistical methods handle extrapolation can have drastic effects on predictions
401 (Pearson et al. 2006).
402 When using complex models, feature space may be sparsely sampled, which means that
403 when one expects to interpolate a predictor, there may be inadvertent extrapolation of
404 nonlinear features. For example, in a model with interaction terms, one may adequately
405 sample the linear features for all predictors while poorly sampling the relevant combinations
406 of these predictors (Zurell et al. 2012). Complex models can lead to different combinations of
407 features producing similar model performance in the present (Maggini et al. 2006), but vastly
408 diverging spatial predictions when transferred to other conditions (Thuiller 2003, Thuiller et
409 al. 2004, Pearson et al. 2006, Edwards et al. 2006, Elith et al. 2010). Narrowing the range of
410 possibilities using a simpler model that controls for the biological plausibility of the response
411 curves (cf. section I.B) can reduce this divergence.

412 **C. Data attributes**

413 *1. Sample Size*

414 The number of occurrence records is a critical limiting factor when building SDMs. With
415 presence-absence data, the number of records in the least frequent class determines the

amount of information available for modeling. Small sample sizes can lead to low signal to noise ratios, thereby making it difficult to evaluate the strength of any occurrence-environment pattern in the presence of confounding processes.

Recommendations: Simple models are necessary for species with few occurrences to avoid over-fitting (Fig. 1). This suggests few predictors and only simple features. Support for features can be found by reporting intervals on response curves (e.g. from confidence intervals or subsamples), with an eye for tight intervals around pronounced nonlinearities. For large data sets, any of the modeling paradigms discussed here are suitable.

Simple: We expect a large amount of noise in occurrence data due to processes unrelated to environmental responses and this noise can be particularly influential when sample sizes are small. For example, if a basic temperature response is built from data that are variably influenced by a strong land-use history and dispersal limitation throughout the range, a failure to take that into account results in a mis-specified climate response surface. Complex models can more readily fit these latent patterns, leading to biased prediction when models are projection to other locations where the latent processes differ. Complex models fitting many features are only appropriate when there are sufficient data to meaningfully train, test and validate the model (cf. Hastie et al. 2009).

Complex: If data are available, increasing the number of predictors ensures a more accurate understanding of the drivers of distributions. If the data set is small, it is possible to use a method that can be potentially complex, as long as it is well controlled by the user to protect against over-fitting e.g., using penalized likelihoods (Tibshirani 1996), a reduced set of features in MAXENT; (Phillips and Dudik 2008, Merow et al. 2013), or heavy pruning in tree-based methods. Permitting some complexity may be useful to identify counterintuitive response curves and develop stratified sampling strategies for future data collection to support or refute the model responses.

441 *2. Sampling Bias*

442 Sampling bias arises from imperfect sampling design, which includes purposive, non-
443 probabilistic, or targeted sampling (Schreuder et al. 2001, Edwards et al. 2006) and imperfect
444 detection (MacKenzie et al. 2002). The important question is whether sampling bias – which
445 often arises in geographic space – transfers to bias in environmental space, and further,
446 whether some environments are completely unsampled. No statistical manipulation can fully
447 overcome biased sampling. The main challenge when choosing complexity is that –
448 particularly for models based on presence-only data – it may be unclear whether patterns in
449 environmental space derive from habitat suitability, divergence between the fundamental and
450 realized niches (Pulliam 2001), or sampling problems (Phillips et al. 2009, Warton et al.
451 2013, Hefley et al. 2013). For presence-absence data with perfect detection, sampling biases
452 may not be too detrimental as long as at least some samples exist across environments into
453 which the model is required to predict (Zadrozny 2004; but see Edwards et al 2006 for
454 contrasting results).

455 ***Recommendations:*** More flexible models will be more prone to finding patterns in
456 restricted parts of environmental space where sampling is problematic. Null models can be
457 used to identify spurious predictive power derived from sampling bias (Raes and ter Steege
458 2007). Poor performance on test data could identify over fitting to sampling bias, but only if
459 the test data are unbiased. In practice, if unbiased testing data were available, they could be
460 used to build an unbiased model in the first place. Recent advances that enable presence-only
461 and presence-absence data to be modeled together, and across species, will be useful in this
462 context (Fithian et al. 2014). A tradeoff exists between a complex model that might fit, e.g.,
463 step functions to few data points in poorly sampled regions and simple models that predict
464 smooth but potentially meaningless functions from just a few points.

Simple: The hope when using simple models for biased data is that main trends are still identified. Complex models can over-fit to the bias (particularly if the bias is heterogeneous in space) and miss the true main trends. Methods for dealing with imperfect detection (MacKenzie and Royle 2005, Welsh et al. 2013) or sampling design often specify relatively simple responses to environment because they simultaneously fit the model for sampling (e.g., Latimer et al. 2006), and identifiability can become an issue when too many parameters are used that might relate to either observation or occurrence. In such cases, inference will be limited to very general trends.

Complex: If the sampling bias is strongly linked to the environmental gradients, even simple models can predict spurious relationships (Lahoz-Monfort et al. 2013). Complex models could be useful in understanding, or hypothesizing about, the nature of the sampling bias: for example, the most parsimonious explanation for sharp changes in the probability of presence in some circumstances could be sampling bias, although we know of no published examples. Detection and sampling bias models are not restricted to simple models – for instance, the former have recently been developed for boosted regression trees (Hutchinson et al. 2011) and the latter are often used with MAXENT (Phillips et al. 2009).

3. Predictor Variables: Proximal vs. Distal

A priority in selecting candidate predictors is to identify variables that are as proximal as possible to the factors constraining the species' distribution. Proximal variables (e.g. soil moisture for plants) best represent the resources and direct gradients that influence species ranges (Austin 2002). More distal predictors, such as using topographic aspect as a surrogate for soil moisture, do not directly affect species distributions but do so indirectly through their imperfect relationships with the proximal predictors they replace. The problem with using distal predictors is that their correlation with the proximal predictor can change across the species' range, even if the proximal predictor's relationship with the species does not

(Dormann et al. 2013). We rarely have access to all of the most important proximal predictors across a study region, so the main question is what response shapes should we expect for more distal predictors? Imagine that a species is limited by the duration of the growing season, but that the response is instead modeled with a combination of mean annual temperature and topographic position (aspect, slope, etc.). It is difficult to anticipate the shape of the multivariate surface that mimics the species response to the proximal predictor.

Recommendations: Responses to proximal predictors over sufficiently large gradients should be relatively strong (Austin, 2007 and references therein), and either simple or complex models should be able to identify these responses if complexity is suitably controlled. However, the extent to which the included set of predictors is proximal or distal may be unknown. Experimentation with complex and simple models may help test hypotheses about which predictors are more proximal, potentially best encapsulated in a simple response curve, and those that are more distal and better represented with more complex curves. As physiological mechanisms generally provide the best insights into how environmental gradients translate into demographic (and therefore population) patterns, using informed physiological understanding could provide a valuable starting point (Austin 2007, Kearney and Porter 2009).

Simple: Ecological theory supports using unimodal or skewed smooth responses to proximal variables (Oksanen 1997, Austin and Nicholls 1997, Austin 2002, Guisan and Thuiller 2005, Austin 2007, Franklin 2010), which motivates constraining the functional form of response curves *a priori* (section I.B; e.g. specific features in a GLM, few nodes in a GAM). Remotely sensed data, even for proximal predictors, may introduce noise to the environmental covariates due to imprecision and to use of long term averaged data (Austin 2007 and references therein, Letten et al. 2013), and may be prone to over-fitting with complex models if those data generally fail to describe the local habitat conditions accurately.

One can use simple models can smooth over such idiosyncrasies if the main trends are sufficiently strong or omit predictors if trends are weak. Parametric, latent variable models can help to deal with this imprecision (McInerny and Purves 2011).

Complex: Ecological theory is based on responses to idealized gradients, whereas we observe (often imperfectly) a messy reality. Complex models work on the assumption that we should not necessarily expect to see theoretical relationships in any given dataset. Specifying an overly simple model will result in over- and under-estimation of the response at points throughout the covariate space (Barry and Elith 2006). Given that the relationship between proximal and distal predictors is unlikely linear and may vary across landscapes, it is likely that the true response to distal variables might also be complex and best represented by a model that allows flexible fits and interactions. Hence the complex viewpoint still adheres to ecological theory, but allows for a modified view of idealized relationships as seen through available data.

4. Spatial Extents and Resolution

Interpretation of ecological patterns is scale dependent; hence changing spatial extent and/or resolution affects the patterns and processes that can be modeled (Tobalske 2002, Chave 2013). Ecologists often use hierarchical concepts to describe influences of environment on species distributions – for instance, that climate dominates distributions of terrestrial species at the global scale (coarsest grain, largest extent), while topography, lithology or habitat structure create the finer scale variation that impact species at regional to local scales together with dispersal limitations and biotic interactions (Dubuis et al. 2012, Boulangeat et al. 2012, Thuiller et al. 2013). SDMs built across large spatial extents often rely on remotely sensed, coarse resolution or highly interpolated predictors, creating inherent biases and sampling issues (section II.C.2). The choice of extent can also determine whether

539 the species entire range is included in the model or whether data are censored (e.g., limited by
540 political borders).
541 ***Recommendations:*** Resolutions should be chosen that provide data from proximal rather
542 than distal variables. Such data are becoming available at high resolutions with expanded and
543 technologically enhanced monitoring networks and more sophisticated interpolation of
544 climate data (e.g., PRISM). The choice of resolution hence reduces to the discussion of
545 proximal versus distal predictors in Section II.C.3. When the extent is chosen to contain the
546 species' entire range, models should include sufficient complexity to detect unimodal,
547 skewed responses (section I.B).
548 ***Simple:*** Smooth responses, characterized by simpler models, are to be expected at large
549 spatial extents and coarse resolution that smooth over the confounding processes that affect
550 finer resolution occurrence patterns (Austin 2007). At finer resolutions, it may also be
551 undesirable to incorporate the full complexity of the response curve: much of the finer details
552 may derive from factors for which no predictor variables are available or are irrelevant to the
553 purpose of the investigation (e.g. microhabitat or regional competition effects).
554 ***Complex****:* At small spatial extents, we might have data on the relevant proximal factors
555 (e.g. soil properties), so fitting complex models along small-scale gradients can capture this
556 complexity. Also, complex models may be useful for exploring the nonlinearities that arise in
557 response curves from distal variables at broad scales in that they potentially provide insight
558 into important unmeasured variables.

559 *5. Spatial Autocorrelation*

560 Many processes omitted from SDMs have spatial structure. For example, dispersal
561 limitation, foraging behavior, competition, prevailing weather patterns, and even sampling
562 bias can all lead to spatially structured occurrence patterns that are not explained by the set of
563 predictors included in the SDM (Legendre 1993, Barry and Elith 2006 but see Latimer et al

2006; Dormann et al. 2007). When these spatial patterns are not appropriately accounted for, biased estimates of environmental responses may emerge.

Recommendations: If presence-absence data are available, assess the degree of spatial autocorrelation in the residuals and implement methods to control for spatial autocorrelation inadvertently influencing the response curves. Methods include spatially-explicit models that separate the spatial pattern from the environmental response (Latimer et al. 2006, Dormann et al. 2007, Beale et al. 2010), using spatial eigenvectors as predictors (Diniz-Filho and Bini 2005), or stratified sub-sampling of the data to minimize autocorrelation (Hijmans 2012). Complex models should be used cautiously in the presence of spatial autocorrelation, because their flexibility may lead to them confounding aggregation in geographic space with complexity in environmental space. For example, if a large number of presences are recorded in a small region of environmental space due to social behavior in geographic space, it is more likely that a complex model can find some feature in environmental space that correlates with this clustering. This will result in biased interpretation or mapped projections in other locations where this social behavior is absent. Cross-validation can eliminate such spurious fits, but only if it is spatially stratified at an appropriate scale. However, when used for exploratory purposes, complex models may reveal information about this spatial structure within their response curves.

Simple: Simple parametric models can accommodate spatial structure under assumptions about the correlation structure (Latimer et al 2006; Dormann et al. 2007). If a non-spatial model is used, simple models can be valuable because they are not flexible enough to model discontinuities in the response curve that derive from spatial structure, however they will still exhibit bias due to aggregated observations. Another solution to dealing with spatial aggregation is to model at sufficiently coarse resolution (suggesting simple models; see *Spatial Extents and* Resolution) that geographic clustering occurs within (and not among)

589 cells, so it can effectively be ignored. One should be cautious building complex models
590 because in practice, obtaining spatially independent cross-validation samples is extremely
591 challenging when the underlying spatial process is unknown and failing to do so likely leads
592 to over-fitting (cf. Hijmans 2012).

593 ***Complex****:* It may be desirable to use complex response curves as proxies for geographic
594 clustering for mapping applications if the model focuses on small extents where nonlinear
595 relationships are likely to hold across the landscape of interest (e.g., interpolation). For
596 example, Santika and Hutchinson (2009) showed that using only linear responses in logistic
597 regression reduced the model performance by misleadingly introducing spatial
598 autocorrelation in the residuals, instead of allowing for unimodal responses in semi-
599 parametric GAMs. Models broadly dealing with spatial and temporal autocorrelation are
600 more recently available for complex models (Hothorn et al. 2011, Crase et al. 2012).

601 **III. CONCLUSIONS**

602 *A. Methodological*

603 Based on our observations on the appropriate use of different statistical methods and
604 modeling decisions, how should modelers proceed to build SDMs? Many modelers'
605 preferences for particular statistical methods derive from the types of data they typically use
606 and the questions they ask, rather than any fundamental philosophy of statistical modeling.
607 For this reason, it is valuable for modelers to have experience in both simple and complex
608 modeling strategies. We suggest that researchers develop a comprehensive understanding of
609 regression models in general and GLMs in particular, as these represent the foundation of
610 almost all of the more complex modeling frameworks. Also, understanding at least one
611 approach to building complex SDMs can allow for sequential tests of more complex model
612 structure. Importantly, because there are many different approaches to handling the same

613 challenges in the data, it is less critical to understand each and every modeling paradigm than
614 to become an expert in applying representatives of simple and complex modeling approaches.
615 Bias can come from over fitting complex models, and it can come from mis-specified
616 simple models. To find a model of optimal complexity, many approaches are possible and
617 are readily justified if sufficient cross-validation has been performed. One might consider
618 starting simple and adding the minimum complexity necessary, or conversely starting with a
619 complex model and removing as much superfluous complexity as possible. If one can narrow
620 down the potential complexity based on the considerations discussed here to consider models
621 within a particular modeling paradigm (Table 1), then traditional model selection techniques
622 are appropriate (section I.B.2).
623 Due to the exploratory nature of many SDMs and the desire to discover spatial patterns
624 and their drivers, we recommend that analyses begin exploration using complex models to
625 determine an upper bound on the complexity of response curves. Over fitting can be
626 controlled through cross-validation (e.g. k-fold, and particularly block resampling methods),
627 even if a full decomposition of into train-validation-test data is not feasible. Furthermore,
628 complex models can be used to identify smooth, simple occurrence-environment
629 relationships if patterns are sufficiently strong and guide specification of simpler models. In
630 contrast, it will be more difficult to overcome a mis-specified simple model, should a more
631 complex response exist. If the exploration with complex models reveals smooth relationships,
632 one can shift to a simpler model. If instead strong nonlinearities are prevalent, consider
633 biological explanations for the nonlinearities. If complex nonlinearities cannot be avoided,
634 focus on minimizing the complexity, understanding it through sensitivity analysis and
635 uncertainty analysis (see below) and providing biologically based justification or hypotheses
636 about it. The end result is a model that adds complexity only to the extent necessary to
637 reproduce observed patterns.

638 Uncertainty analysis is a relatively untapped resource for understanding appropriate
639 model complexity. When the influence of particular model components is unknown (e.g.
640 whether a predictor or feature is relevant *a priori*) it is particularly critical to account for
641 uncertainty in modeled relationships to explore the implications of our ignorance. By
642 studying uncertainty, one can gain confidence in pronounced nonlinearities when they come
643 with tight confidence intervals. Information on parameter uncertainty, and consequently
644 prediction uncertainty, can be obtained from any means of simulation from parameter
645 distributions, including posterior sampling, sampling based on point estimates and covariance
646 matrices, or bootstrapping. Bayesian models have the advantage of using the full data set to
647 estimate parameter uncertainty, but are generally restricted to simpler models to avoid
648 convergence issues (e.g., Latimer et al. 2006, Ibáñez et al. 2009). One way of reducing
649 uncertainty in predictions is to analyze the importance of predictors given the model and data
650 using 'average predictive comparisons' (Gelman and Pardoe 2007) a form of sensitivity
651 analysis that incorporates parameter uncertainty. One can also quantify uncertainty due to our
652 modeling decisions by using ensembles of models built with different statistical methods or
653 decisions (Pearson et al. 2006, Araújo and New 2007, Thuiller et al. 2009), provided that
654 each component model is built based on modeling decisions reflecting a common goal.

655 *B. Biological*

656 Despite the valuable insights we can gain from occurrence models, it is worth acknowledging
657 that fundamental limitations to biological inference may emerge from these studies (Tyre et
658 al. 2001, Araújo and Guisan 2006, Araújo and Peterson 2012, Merow et al. 2013). Balancing
659 complex and simple models in such a way as to discover and discuss these limits may be as
660 important as the actual patterns identified with some datasets. More broadly, it is important to
661 keep in mind that we are ultimately performing exploratory analyses of occurrence-
662 environment relationships. Occurrence records are not the ideal data to predict attributes of

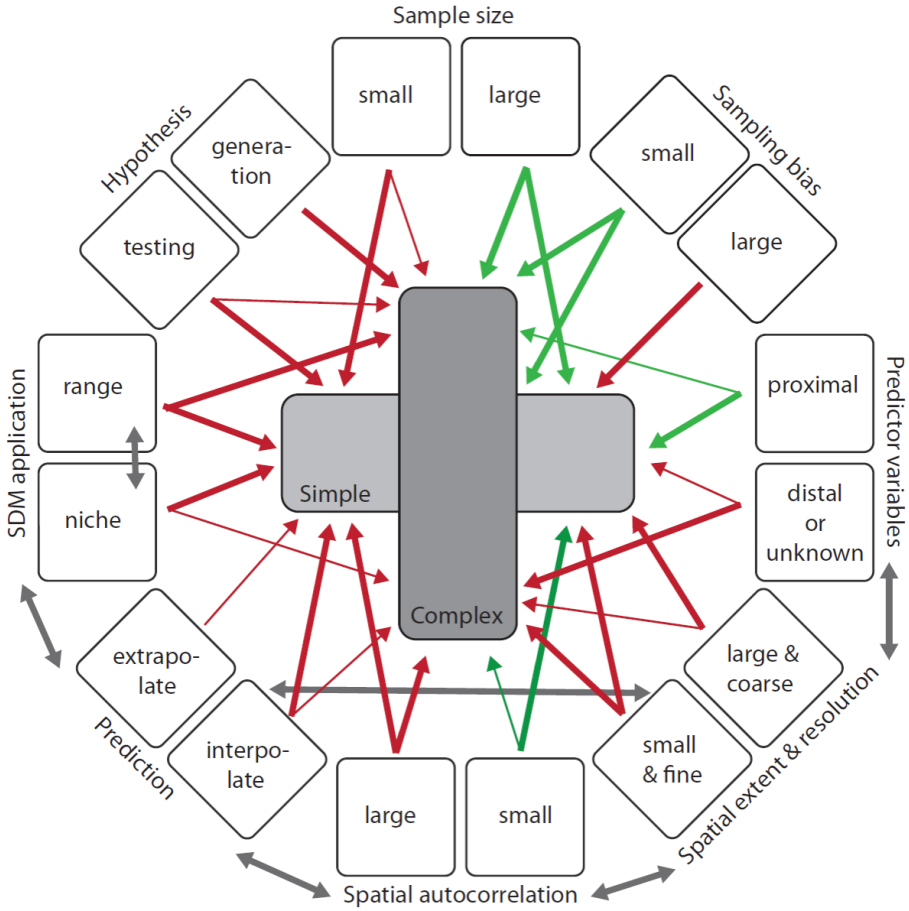
populations, but often no other data are available at large spatial extents that might inform range models. Thus, while the limits may be obvious, insights from occurrence-based correlative models may be an essential step in developing new hypotheses and research programs that can lead to the next generation of mechanistic models (Thuiller et al. 2013; Schurr et al. 2012).

A novel, and potentially important, application of SDMs is for informing mechanistic models about the shapes of response curve in demographic models (Edwards et al., this issue, Merow et al., this issue), or dynamic spatio-temporal population models (Pagel and Schurr 2012, Thuiller et al., this issue; Boulangeat et al., this issue). Simple models may be preferable for these tasks because it is important to have a clear hypothesis to evaluate when linking it to a particular process (Thuiller et al. 2013). For example, SDMs might inform variable selection for the growth, survival and fecundity models in Integral Projection Models (Easterling et al. 2000). However highly nonlinear relationships would not be desirable for vital rate models due to the unlikely transitions through the life history that they might imply (cf. Merow et al., this issue). It is particularly important to avoid confounding missing processes with complex environmental responses (as might occur in complex models) when the mechanistic model explicitly describes the mechanisms that produce that aggregation (e.g. dispersal or species interactions: Kissling et al. 2012). The challenge in using SDMs in this way lies in ensuring response curves truly reflect environmental limitations; while environmental tolerance may limit a species' distribution at one end of a gradient, other (e.g. biotic) factors may limit it at the other end (Zimmermann et al. 2009). Progress towards improved process-based models lies in challenging occurrence-based SDMs with stronger biological justifications and interpretations that aim to detect the true underlying complexity in biological responses.

ACKNOWLEDGEMENTS

This study arose from two workshops entitled ‘Advancing concepts and models of species range dynamics: understanding and disentangling processes across scales’. Funding was provided by the Danish Council for Independent Research | Natural Sciences (grant no. 10-085056 to SN). CM acknowledges funding from NSF grant 1046328 and NSF grant 1137366. WT acknowledges support from the European Research Council under the European Community's Seven Framework Programme FP7/2007-2013 Grant Agreement no. 281422 (TEEMBIO). RW acknowledges support from the Swiss National Science Foundation (Synergia Project CRS113-125240, Early Postdoc .Mobility Grant PBZHP3_147226). JE acknowledges funding from the Australian Research Council (grant FT0991640). TE states that mention of any product by names does not constitute endorsement by the U.S. Geological Survey.

Figure 1



Influence of attributes of study objectives and data attributes on the choice of model complexity. Green arrows illustrate attributes where the choice of complexity is of no particular concern. Red arrows illustrate the situations where caution and/or experimentation with model complexity is needed. Gray arrows indicate decisions that involve interactions with other study goals or data attributes. The thickness of the arrows illustrates the strength of the arguments in favor of choosing a specific level of complexity, with thicker arrows indicating stronger arguments.

IV. References

Aarts, G. et al. 2012. Comparative interpretation of count, presence-absence and point methods for species distribution models. - Methods in Ecology and Evolution 3: 177–187.

Albert, C. H. et al. 2010. A multi-trait approach reveals the structure and the relative importance of intra- vs. interspecific variability in plant traits. - Functional Ecology 24: 1192–1201.

Araujo, M. B. and Peterson, A. T. 2012. Uses and misuses of bioclimatic envelope modelling. - Ecology 93: 1527–1539.

Araújo, M. and New, M. 2007. Ensemble forecasting of species distributions. - Trends in Ecology and Evolution 22: 42–47.

Araújo, M. B. and Guisan, A. 2006. Five (or so) challenges for species distribution modelling. - J Biogeography 33: 1677–1688.

Araújo, M. B. et al. 2005. Validation of species–climate impact models under climate change. - Global Change Biol 11: 1504–1513.

Austin, M. 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. - Ecological Modelling 200: 1–19.

Austin, M. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. - Ecological Modelling 157: 101–118.

Austin, M. P. 1976. On Non-Linear Species Response Models in Ordination. - Vegetatio 33: 33–41.

734 Austin, M. P. and Nicholls, A. O. 1997. To fix or not to fix the species limits, that is the
735 ecological question: Response to Jari Oksanen. - Journal of Vegetation Science 8: 743–
736 748.

737 Austin, M. P. and Smith, T. M. 1989. A New Model for the Continuum Concept. - Vegetatio
738 83: 35–47.

739 Austin, M. P. et al. 1994. Determining species response functions to an environmental
740 gradient by means of a β-function. - Journal of Vegetation Science 5: 215–228.

741 Barry, S. and Elith, J. 2006. Error and uncertainty in habitat models. - Journal of Applied
742 Ecology 43: 413–423.

743 Beale, C. M. et al. 2010. Regression analysis of spatial data. - Ecol Letters 13: 246–264.

744 Boulangeat, I. et al. 2012. Accounting for dispersal and biotic interactions to disentangle the
745 drivers of species distributions and their abundances. - Ecol Letters 15: 584–593.

746 Breiman, L. 2001. Statistical modeling: the two cultures. - Statist. Sci. 16: 199–231.

747 Breiman, L. et al. 1984. Classification and Regression Trees. - Wadsworth International
748 Group.

749 Buckley, L. B. et al. 2011. Does including physiology improve species distribution model
750 predictions of responses to recent climate change? - Ecology 92: 2214–2221.

751 Burnham, K. and Anderson, D. R. 2002. Model Selection and Multimodel Inference: A
752 Practical Information-Theoretic Approach. - Springer-Verlag.

753 Chase, J. M. and Leibold, M. A. 2003. Ecological Niches. - University of Chicago Press.

754 Chave, J. 2013. The problem of pattern and scale in ecology: what have we learned in

20 years? - Ecol Letters 16: 4–16.

Crase, B. et al. 2012. A new method for dealing with residual spatial autocorrelation in species distribution models. - Ecography 35: 879–888.

Diniz-Filho, J. and Bini, L. M. 2005. Modelling geographical patterns in species richness using eigenvector-based spatial filters. - Global Ecol Biogeography 14: 177–185.

Dormann, C. F. et al. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. - Ecography 36: 27–46.

Dormann, C. F. et al. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. - Ecography 30: 609–628.

Dubuis, A. et al. 2012. Improving the prediction of plant species distribution and community composition by adding edaphic to topo-climatic variables. - Journal of Vegetation Science 24: 593–606.

Easterling, M. R. et al. 2000. Size-specific sensitivity: applying a new structured population model. - Ecology 81: 694–708.

Edwards, T. C., Jr et al. 2006. Effects of sample survey design on the accuracy of classification tree models in species distribution models. - Ecological Modelling 199: 132–141.

Elder, J. F., IV 2003. The Generalization Paradox of Ensembles. - Journal of Computational and Graphical Statistics 12: 853–864.

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

776 77.

777 Elith, J. and Leathwick, J. R. 2009. Species Distribution Models: Ecological Explanation and
778 Prediction Across Space and Time. - Annu. Rev. Ecol. Evol. Syst. 40: 677–697.

779 Elith, J. et al. 2005. The evaluation strip: a new and robust method for plotting predicted
780 responses from species distribution models. - Ecological Modelling 186: 280–289.

781 Elith, J. et al. 2010. The art of modelling range-shifting species. - Methods in Ecology and
782 Evolution 1: 330–342.

783 Fithian, W. et al. 2014. A Proportional Observer Bias Model for Multispecies Distribution
784 Modeling. - arXiv in press.

785 Fox, J. 2003. Effect displays in R for generalised linear models. - Journal of Statistical
786 Software 8: 1–27.

787 Franklin, J. 2010. Moving beyond static species distribution models in support of
788 conservation biogeography. - Diversity and Distributions 16: 321–330.

789 Friedman, J. H. 2001. Greedy function approximation: a gradient boosting machine. - The
790 Annals of Statistics: 1189–1232.

791 Gelman, A. and Pardoe, I. 2007. Average predictive comparisons for models with
792 nonlinearity, interactions, and variance components. - Sociological Methodology 37: 23–
793 51.

794 Grömping, U. 2009. Variable Importance Assessment in Regression: Linear Regression
795 versus Random Forest. - The American statistician 63: 308–319.

796 Guisan, A. and Thuiller, W. 2005. Predicting species distribution: offering more than simple

habitat models. - Ecol Letters 8: 993–1009.

Guisan, A. et al. 2013. Predicting species distributions for conservation decisions. - Ecology 16: 1424–1435.

Hastie, T. et al. 2009. The elements of statistical learning: data mining, inference, and prediction. - Springer-Verlag.

Hastie, T. J. and Tibshirani, R. J. 1990. Generalized Additive Models. - Chapman Hall in press.

Hefley, T. J. et al. 2013. Nondetection sampling bias in marked presence-only data. - Ecol Evol: 1–12.

Hijmans, R. J. 2012. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. - Ecology 93: 679–688.

Hirzel, A. et al. 2006. Evaluating the ability of habitat suitability models to predict species presences. - Ecological Modelling 199: 142–152.

Holt, R. D. 2009. Bringing the Hutchinsonian Niche into the 21st Century: Ecological and Evolutionary Perspectives. - P Natl Acad Sci Usa 106: 19659–19665.

Hothorn, T. et al. 2011. Decomposing environmental, spatial, and spatiotemporal components of species distributions. - Ecological Monographs 81: 329–347.

Hutchinson, R. A. et al. 2011. Incorporating Boosted Regression Trees into Ecological Latent Variable Models. - Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence: 1343–1348.

Ibáñez, I. et al. 2009. Multivariate forecasts of potential distributions of invasive plant

818 species. - Ecol Appl 19: 359–375.

819 Janson, L. et al. 2013. Effective Degrees of Freedom: A Flawed Metaphor. - arXiv in press.

820 Kearney, M. and Porter, W. 2009. Mechanistic niche modelling: combining physiological
821 and spatial data to predict species' ranges. - Ecol Letters 12: 334–350.

822 Kissling, W. D. et al. 2012. Towards novel approaches to modelling biotic interactions in
823 multispecies assemblages at large spatial extents. - J Biogeography 39: 2163–2178.

824 Lahoz-Monfort, J. J. et al. 2013. Imperfect detection impacts the performance of species
825 distribution models. - Global Ecol Biogeography 23: 504–515.

826 Latimer, A. M. et al. 2006. Building statistical models to analyze species distributions. - Ecol
827 Appl 16: 33–50.

828 Lawson, C. R. et al. 2013. Prevalence, thresholds and the performance of presence-absence
829 models. - Methods in Ecology and Evolution 5: 54–64.

830 Legendre, P. 1993. Spatial Autocorrelation: Trouble or New Paradigm? - Ecology 74: 1659–
831 1673.

832 Letten, A. D. et al. 2013. The importance of temporal climate variability for spatial patterns
833 in plant diversity. - Ecography 36: 1341-1349.

834 Lobo, J. et al. 2008. AUC: a misleading measure of the performance of predictive distribution
835 models. - Global Ecol Biogeography 17: 145–151.

836 MacKenzie, D. I. and Royle, J. A. 2005. Designing occupancy studies: general advice and
837 allocating survey effort. - Journal of Applied Ecology 42: 1105–1114.

838 MacKenzie, D. I. et al. 2002. Estimating site occupancy rates when detection probabilities are

839 less than one. - Ecology 83: 2248–2255.

840 Maggini, R. et al. 2006. Improving generalized regression analysis for the spatial prediction
841 of forest communities. - J Biogeography 33: 1729–1749.

842 Mcinerny, G. J. and Purves, D. W. 2011. Fine-scale environmental variation in species
843 distribution modelling: regression dilution, latent variables and neighbourly advice. -
844 Methods in Ecology and Evolution 2: 248–257.

845 Merow, C. et al. 2013. A practical guide to MaxEnt for modeling species' distributions: what
846 it does, and why inputs and settings matter. - Ecography 36: 1–12.

847 Normand, S. et al. 2009. Importance of abiotic stress as a range-limit determinant for
848 European plants: insights from species responses to climatic gradients. - Global Ecol
849 Biogeography 18: 437–449.

850 Oksanen, J. 1997. Why the beta-function cannot be used to estimate skewness of species
851 responses. - Journal of Vegetation Science 8: 147–152.

852 Pearce, J. and Ferrier, S. 2000. Evaluating the predictive performance of habitat models
853 developed using logistic regression. - Ecological Modelling 133: 225–245.

854 Pearson, R. G. et al. 2013. Shifts in Arctic vegetation and associated feedbacks under climate
855 change. - Nature Climate change in press.

856 Pearson, R. G. et al. 2006. Model-based uncertainty in species range prediction. - J
857 Biogeography 33: 1704–1711.

858 Peterson, A. T. et al. 2011. Ecological Niches and Geographic Distributions. - Princeton
859 University Press.

860 Phillips, S. and Dudik, M. 2008. Modeling of species distributions with Maxent: new
861 extensions and a comprehensive evaluation. - Ecography 31: 161–175.

862 Phillips, S. et al. 2006. Maximum entropy modeling of species geographic distributions. -
863 Ecological Modelling 190: 231–259.

864 Phillips, S. et al. 2009. Sample selection bias and presence-only distribution models:
865 implications for background and pseudo-absence data. - Ecological Applications 19:
866 181–197.

867 Pulliam, H. R. 2000. On the relationship between niche and distribution. - Ecol Letters 3:
868 349–361.

869 Raes, N. and terSteege, H. 2007. A null-model for significance testing of presence-only
870 species distribution models. - Ecography 30: 727–736.

871 Randin, C. F. et al. 2006. Are niche-based species distribution models transferable in space? -
872 J Biogeography 33: 1689–1703.

873 Santika, T. and Hutchinson, M. F. 2009. The effect of species response form on species
874 distribution model prediction and inference. - Ecological Modelling 220: 2365–2379.

875 Schreuder, H. T. et al. 2001. For what applications can probability and non-probability
876 sampling be used? - Environmental Monitoring and Assessment 66: 281–291.

877 Soberón, J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. -
878 Ecol Letters 10: 1115–1123.

879 Strobl, C. et al. 2008. Conditional Variable Importance for Random Forests. - BMC
880 Bioinformatics 9: 307.

881 Svenning, J.-C. et al. 2011. Applications of species distribution modeling to paleobiology. -
882 Quaternary Science Reviews 30: 2930–2947.

883 Thuiller, W. 2003. BIOMOD–optimizing predictions of species distributions and projecting
884 potential future shifts under global change. - Global Change Biol 9: 1353–1362.

885 Thuiller, W. et al. 2009. BIOMOD–A platform for ensemble forecasting of species
886 distributions. - Ecography 32: 369–373.

887 Thuiller, W. et al. 2004. Effects of restricting environmental range of data to project current
888 and future species distributions. - Ecography 27: 165–172.

889 Thuiller, W. et al. 2013. A road map for integrating eco-evolutionary processes into
890 biodiversity models. - Ecol Letters 16: 94–105.

891 Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. - Journal of the Royal
892 Statistical Society. Series B (Methodological) 58: 267–288.

893 Tobalske, C. 2002. Effects of spatial scale on the predictive ability of habitat models for the
894 Green Woodpecker in Switzerland. - In: Scott, J. M. and al, E. (eds), Predicting Species
895 Occurrences: Issues of Accuracy and Scale. Island Press, ppp. 197–204.

896 Tyre, A. J. et al. 2001. Inferring process from pattern: can territory occupancy provide
897 information about life history parameters? - Ecol Appl 11: 1722–1737.

898 Veloz, S. D. 2009. Spatially autocorrelated sampling falsely inflates measures of accuracy for
899 presence-only niche models. - J Biogeography 36: 2290–2299.

900 Warton, D. I. et al. 2013. Model-Based Control of Observer Bias for the Analysis of
901 Presence-Only Data in Ecology. - PLoS ONE 8: e79168.

902 Welsh, A. H. et al. 2013. Fitting and interpreting occupancy models. - PLoS ONE 8: e52015.

903 Wood, S. N. 2006. Generalized Additive Models. - CRC Press.

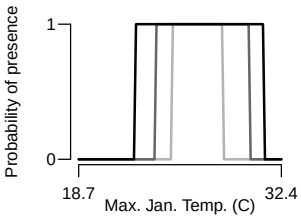
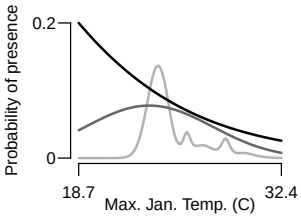
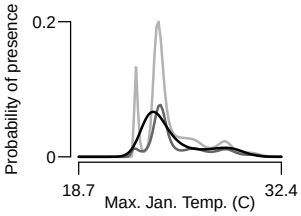
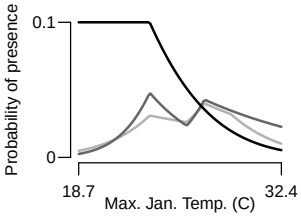
904 Zadrozny, B. 2004. Learning and evaluating classifiers under sample selection bias. - ICML
905 '04: 114.

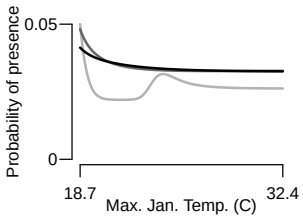
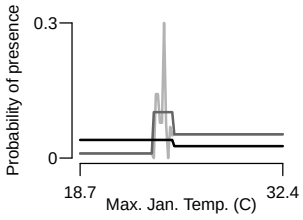
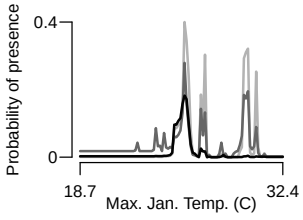
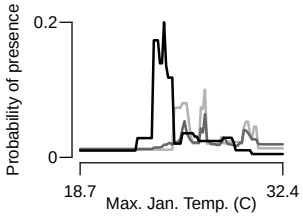
906 Zimmermann, N. E. et al. 2010. New trends in species distribution modelling. - Ecography
907 33: 985–989.

908 Zimmermann, N. E. et al. 2009. Climatic extremes improve predictions of spatial patterns of
909 tree species. - P Natl Acad Sci Usa 106 Suppl 2: 19723–19728.

910 Zurell, D. et al. 2012. Predicting to new environments: tools for visualizing model behaviour
911 and impacts on mapped distributions. - Divers Distrib 18: 628–634.

912

Algorithm	Response Curves	Responses are built from...	Complexity Controlled by...
Bioclimatic Envelope Models (BIOCLIM)		quantiles, between which occurrence probability is 1	• Features: step functions • Quantiles
Generalized Linear Models (GLM)		parametric terms specified by user	• Features: polynomials, piecewise functions, splines • Feature complexity specified by user
Generalized Additive Models (GAM)		combination of parametric terms and flexible smooth functions suggested by the data or the user	• Features: parametric terms as in GLMs and various smoothers (e.g. splines, loess) • Number of nodes • Penalties
Multivariate Adaptive Regression Splines (MARS)		the sum of multiple piecewise basis functions of predictors suggested by the data	• Features: splines • Number of knots • Cost per degree of freedom • Pruning

Algorithm	Response curves	Responses are built from...	Complexity controlled by...
Artificial Neural Networks (ANN)		networks of interactions between simple functions of predictors suggested by the data	• Number of hidden layers
Classification and Regression Trees (CART)		repeated partitioning of predictors into different categories, suggested by the data, associated with different occurrence probabilities	• Features: threshold, with implicit interactions • Minimum observations for split/terminal node • Maximum node depth • Complexity threshold to attempt a split
Random Forests (RF)		an average of multiple CARTs, each constructed on bootstrapped samples of the data and using different random subsets of the full predictor set	• Features: threshold, with implicit interactions • See CARTs • Number of trees
Boosted Regression Trees (BRT)		regression trees at multiple steps; at each, models the residuals from the sum of all previous models weighted by the learning rate	• Features: threshold, with implicit interactions • See CARTs • Number of trees • Learning rate

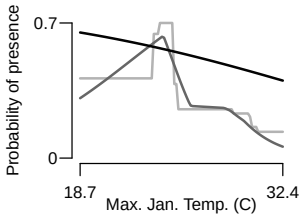
Algorithm	Response Curves	Responses are built from...	Complexity Controlled by...
Maximum Entropy (MAXENT)		a GLM with a large number of features , which are suggested by the data or the user	• Features: linear, quadratic, interaction, hinge, threshold • Feature classes used • Regularization penalty

Table 1: Common modeling paradigms used to build SDMs and decisions used to control their complexity. The variation among response curves from different modeling paradigms and different model settings suggests that they are suitable for different study objectives and attributes of the data. Response curves come from fitting SDMs to presence/background data on the overstory shrub, *Protea punctata*, from the Cape Floristic Region of South Africa (see Merow et al. 2013 for details of the data) with different degrees of control over the complexity of the fitted response curves. All models were constructed using the biomod2 package (Thuiller et al. 2013b) within the statistical software R (R Core Team 2013). Response curves of different complexity are shown which are representative of those commonly observed during SDM building. Dark grey curves were fitted using the settings at or near the default options sets in biomod2 (for illustration) with the exception of forcing the package to perform only a single fit per method using all of the presence data in model fitting. Black (light grey) curves were fitted by choosing options to make the fitted response curves simpler (more complex). Note that complexity of any of these paradigms is affected by changing the number of predictors, the order of interactions, and model averaging, hence these decisions are not explicitly included in the the table.