Review & synthesis

Ecography 37: 1267–1281, 2014
doi: 10.1111/ecog.00845
© 2014 The Authors. Ecography published by John Wiley & Sons Ltd on the behalf of Nordic Society Oikos.
Subject Editor: Heike Lischke. Editor-in-Chief: Jens-Christian Svenning. Accepted 24 July 2014

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

Cory Merow, Mathew J. Smith, Thomas C. Edwards Jr, Antoine Guisan, Sean M. McMahon, Signe Normand, Wilfried Thuiller, Rafael O. Wüest, Niklaus E. Zimmermann and Jane Elith

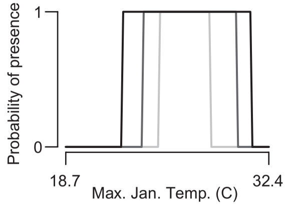
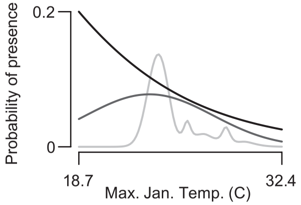
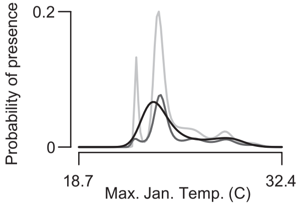
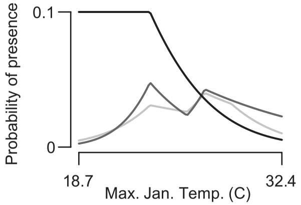
C. Merow (cory.merow@gmail.com) and S. M. McMahon, Smithsonian Environmental Research Center, Edgewater, MD 21307-0028, USA. CM also at: Univ. of Connecticut, Ecology and Evolutionary Biology, 75 North Eagleville Rd. Storrs, CT 06269, USA. – M. J. Smith and CM, Computational Science Laboratory, Microsoft Research, CB1 2FB, UK. – T. C. Edwards, Jr, U.S. Geological Survey, Utah Cooperative Fish and Wildlife Research Unit and Dept of Wildland Resources, Utah State Univ., Logan, UT 84322, USA. – A. Guisan, Dept d'Ecologie et d'Evolution, Univ. de Lausanne, CH-1015 Lausanne, Switzerland, and Inst. of Earth Surface Dynamics, Univ. de Lausanne, CH-1015 Lausanne, Switzerland. – S. Normand, R. O. Wüest and N. E. Zimmermann, Landscape Dynamics, Swiss Federal Research Inst. WSL, Zürcherstr. 111, CH-8903 Birmensdorf, Switzerland. SN also at: Ecoinformatics and Biodiversity, Dept of Bioscience, Aarhus Univ., Ny Munkegade 116, DK-8000 Aarhus C, Denmark. – W. Thuiller and ROW, Univ. Grenoble Alpes, LECA, FR-38000 Grenoble, France, and 10 CNRS, LECA, FR-38000 Grenoble, France. – J. Elith, Centre of Excellence for Biosecurity Risk Analysis, School of Botany, The Univ. of Melbourne, Parkville 3010, Australia.

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. We characterize complexity as the shape of the inferred occurrence–environment relationships and the number of parameters used to describe them, and search for insights into whether additional complexity is informative 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 approaches. 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.

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 approaches (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 bio-

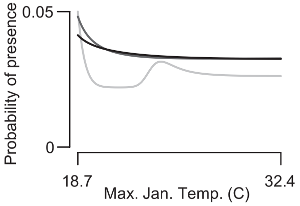
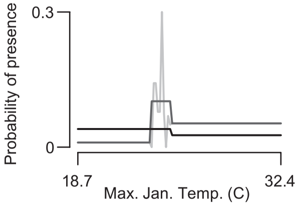
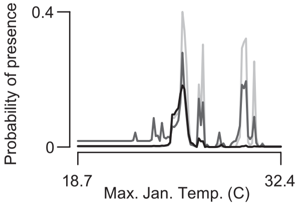
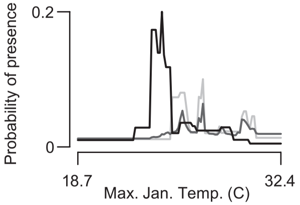
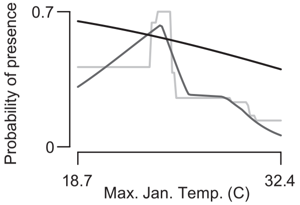
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

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. 2013) within the statistical software R (R Core Team). 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.

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

Continued

Table 1. Continued

Algorithm	Response curves	Responses are built from	Complexity controlled by
Artificial neural networks (ANN)		networks of interactions be tween 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 2	• Features: threshold, with implicit interactions • See CARTs • Number of trees • Learning rate
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

logical processes. Here, we discuss the challenges that choosing an appropriate amount 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 feature of observed occurrence patterns because occurrence–environment relationships may be obscured by processes that are not exclusively related to the environment, such as dispersal, response to disturbance, and biotic interactions (Pulliam 2000, Holt 2009, Boulangeat et al. 2012). Consequently, SDMs can be dynamic and process-based, explicitly representing aspects of the underlying biology. This paper focuses on the more widely used static, correlative SDMs, although many of the issues considered relate to process-based SDMs as well. Describing this complexity is critical for many applications of SDMs, and using flexible occurrence–environment relationships allows biologists to hypothesize about the drivers of complexity or make accurate predictions that derive from their representation in SDMs. Such hypotheses are a valuable step toward the types of process-based models discussed in this issue (Merow et al. 2014, Snell et al. 2014). However, building complex models comes with the challenge of differentiating true complexity from noise (see chapter 7 in Hastie et al. 2009 for a statistical viewpoint on optimising model complexity). Some believe that flexible models are often overfit to the noise prevalent in many occurrence data sets. Thus, with such variation in both needs and opinions regarding model complexity, many modeling approaches are in current use (Table 1).

We characterize model complexity by the shape of the inferred occurrence–environment relationships (Table 1) and the number of posited predictors and parameters used to describe them. A simpler model typically has relatively fewer parameters and fewer relationships among predictors compared to a more complex model. However, it remains a challenge to quantify complexity in a way that is appropriate across the spectrum of modeling approaches in Table 1 (e.g. Janson et al. 2013 showed effective degrees of freedom to be an unreliable metric when defining complexity). Univariate 'response curves' are commonly used to give an impression of the complexity of the predicted occurrence–environment relationships. These are one-dimensional 'slices' of multivariate space. The most common approach is to plot the predicted occurrence probability against the predictor of interest by holding all other predictors at their mean or median values (Elith et al. 2005; Table 1), although other approaches are possible (Fox 2003, Hastie et al. 2009). When visualized in this way, a simpler model is relatively smooth, containing fewer inflection and turning points compared to a more complex model. Though insightful, univariate curves only represent the true fitted response incompletely (3-dimensional response surfaces or the 'inflated response curves' of Zurell et al. (2012) help here). Complex models contain more interactions, which can only be visualized on higher dimensional surfaces, compared to simpler models. Such responses must be interpreted as conditional on the other mean or median predictors in the model, which may be different than the responses to variables held at other values (Zurell et al. 2012), or to an unconditional model. Nonetheless, uni- and multivariate response curves remain one of the best standardized ways to assess relative model complexity.

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 who try to approximate more of the complexities of the real world relationships. We review the literature and the general modeling principles emerging from these two viewpoints, and we 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 approach in order to advance our knowledge of current and future species geographical ranges.

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). One 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, 2002, 2007, Pulliam 2000, Chase and Leibold 2003, Holt 2009). There is, 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 physiological function (Buckley et al. 2011). However, response curves describing realized niches might exhibit discontinuities due to the multiple interacting factors that can limit a species' occurrence in any particular location. Unimodal responses are expected (e.g. a bell-shaped curve) because conditions too extreme for survival often exist at either end of a proximal gradient (Austin 2007). However, response curves can be linear where only part of the environmental range of the species has been sampled (e.g. one side of a unimodal response; Albert et al. 2010). Austin and Smith's (1989) continuum concept for plant species distributions predicts that skewed unimodal response curves are likely when plant species distributions are predominantly determined by one or a few environmental variables that strongly regulate survivorship and or reproduction (e.g. by temperature thresholds), but that more irregular response curves are expected given that species are influenced by a range of regulatory factors (e.g. different limiting nutrients, biotic and abiotic interactions) and historical contingencies (Austin et al. 1994, Normand et al. 2009). Even with single factors, the processes that determine fitness may be different across the range, e.g. where one temperature extreme leads to abrupt loss of function while the other extreme causes gradually reduced performance. Interaction terms can be desirable to capture covariation between predictors or tradeoffs along resource gradients (e.g. higher temperatures are tolerable with greater rainfall). Many applications of SDMs do not explicitly consider such theoretical constraints on the shape of response curves (but see Santika and Hutchinson 2009), perhaps because it is difficult to work out how they translate into observations. We are faced with the challenge of inferring unknown levels of ecological complexity through the lens of data and models that imperfectly capture it.

Complexity in models

Two attributes of model fitting determine the complexity of inferred occurrence–environment relationships in SDMs: the underlying statistical method and modeling decisions made about inputs and settings. Together, these define what we will call different modeling approaches, a number of which are illustrated in Table 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 approaches 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.

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 (see increased complexity of black vs grey MAXENT response curves due to increase in number of features; Table 1). 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, Phillips and Dudik 2008). 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 to account for nonadditive relationships. 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 fitted automatically if supported by the data and cannot be explicitly controlled by the user (except to specify the permissible order of the interactions considered).

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 'multimodel 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 – while these can be used on any component models, in ecology the most used component models are decision trees (e.g. in random forests, Brieman 2001; and boosted regression trees, Friedman 2001). Bagging (bootstrap aggregation) can be used to fit many models to bootstrapped replicates of the dataset (with and without random subsetting of predictors used across trees as in random forests). In contrast, boosting uses a forward stagewise method to build an ensemble, at each step modeling the residuals of the models fitted to date. Taking ensembles of relatively simple models usually increases complexity because combinations of simple models will not necessarily be simple. In contrast, ensembles of more complex models can average over idiosyncrasies of individual models to produce smoother response curves (Elder 2003).

Model comparison

To avoid overfitting and underfitting, it is common to compare models of differing complexity and select the model that optimizes some measure of performance. However, comparing models across modeling approaches (e.g. those in Table 1) can be challenging. This is one of our motivations for constraining model complexity based on study objectives and data attributes. Information theoretic measures are a conventional way to choose model complexity and are relatively easy to apply for models where estimating the number of degrees of freedom is possible. However these cannot be calculated for ensemble-based methods nor for many other methods in common use (Janson et al. 2013). In fact, Janson et al. (2013) warn, 'contrary to folk intuition, model complexity and degrees of freedom are not synonymous and may correspond very poorly'. One way to compare models produced by different algorithms is to adopt a common currency for model performance by evaluating model predictions on either the training data or independent testing data. Measures such as AUC, Cohen's Kappa, and the True Skill Statistic are based on correctly distinguishing presences from absences. Measures based on non-thresholded predictions are also relevant and preferable in many situations (Lawson et al. 2013). However, each of these metrics has weaknesses in different circumstances (Lobo et al. 2008) and further, only represent heuristic diagnostics for presence-only data, because presences must be compared to pseudoabsence/ background data (Hirzel et al. 2006).

Once one has determined a suitable modeling approach tuning of the amount of complexity is more straightforward using a range of model selection techniques. Feature significance (e.g. p-values), measures of model fit (e.g. likelihood), and information criteria (e.g. AIC, AICc, BIC; Burnham and Anderson 2002) can be applied to regression-based methods. Cross-validation or other resampling techniques are also used to set the smoothness of splines in GAMs (Wood 2006) or to determine tuning parameters in most machine learning methods (Hastie et al. 2009). Shrinkage or regularization is often used in regression, MAXENT and boosted regression trees to constrain coefficient estimates so models predict reliably (Phillips et al. 2006, Hastie et al. 2009). Loss functions, which penalize for errors in prediction, can be constructed for any of the modeling approaches we consider (Hastie et al. 2009). An alternative approach employs null models to evaluate whether additional complexity has lead to spurious predictive accuracy (Raes and terSteege 2007).

Evaluation against fit to training data alone cannot control for over fitting and risks selecting excessively complex models (Pearce and Ferrier 2000, Araújo et al. 2005). In general, best practice involves splitting the data into training data to fit the model, validation data for model selection, and test data to evaluate the predictive performance of the selected model (Hastie et al. 2009). Recent studies have emphasized that care should also be taken in how data is partitioned into training, evaluation and test data, in particular to control for spatial autocorrelation (Latimer et al. 2006, Dormann et al. 2007, Veloz 2009, Hijmans 2012; see below for more details). Hence methods such as block cross-validation (where blocks are spatially stratified) are gaining momentum (Hutchinson et al. 2011, Pearson et al. 2013, Warton et al. 2013). Failure to factor out spatial autocorrelation in data 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 below).

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 approach, but to benefit from these, it is valuable to narrow the scope of the feasible models based on biological considerations. We therefore now move to exploring approaches for identifying the appropriate level of complexity for particular study objectives based on data limitations and the underlying biological processes.

Philosophical, statistical and biological considerations when choosing complexity

In this section, we discuss factors that should influence the choice of model complexity. First, we outline general considerations and philosophical differences underlying both simple and complex modeling strategies (section Simple versus complex: fundamental approaches to describing natural systems). Next, we discuss how the study goals (section Study objectives) and data attributes (section Data attributes) interact with model complexity. Figure 1 summarizes our findings. Importantly, a general consensus for choosing model complexity is not possible in many cases. To reflect the different schools of thought, we divide our facts, ideas and opinions into those that are relatively uncontroversial (subsections denoted 'Recommendations'), those that favor simple models (denoted 'Simple'), and those that favor more complex models (denoted 'Complex'). We recall that 'simple' and 'complex' refer to the extremes along a gradient of complexity in response curves pro-

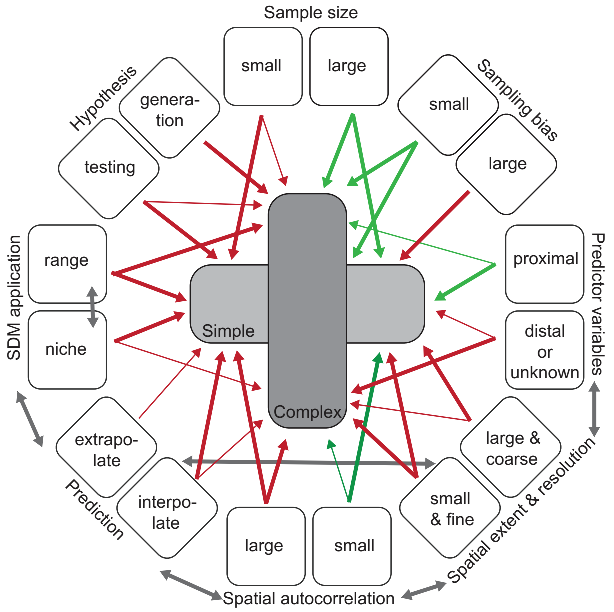


Figure 1. Influence of attributes of study objectives and data attributes on the choice of model complexity. Green arrows illustrate attributes where the choice of complexity is of no particular concern. Red arrows illustrate the situations where caution and/or experimentation with model complexity is 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.

duced by distinct statistical methods and modeling decisions (section Complexity in models and Table 1).

Simple versus complex: fundamental approaches to describing natural systems

Simple

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

Complex

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

Study objectives

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 affect 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 Interpolate vs extrapolate), 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 Spatial autocorrelation).

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 Complexity in models; 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 (Elith et al. 2005, Randin 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.

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. For example, SDMs are valuable exploratory analyses for detecting the processes that confound occurrence–environment relationships, such as transient dynamics, dispersal, biotic interactions, or human modification of landscapes. The indirect effect of such processes can be seen in occurrence patterns, often due to abrupt changes or nonlinearities in response curves, leading to hypothesis generation. Whether one is testing or generating hypotheses 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 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.

Simple

When the goal is hypothesis testing, simple parametric models allow investigation of the strength and shape of relationships between species occurrence and a small set of features. Furthermore, parametric models allow for hypothesis tests to examine if specific nonlinear features should be included in the selected model(s). The problem with complex models in such a setting is that with the large suite of potential features that they use, it is challenging to determine the significance of a single feature or attribute of the response curve or to compare alternative models. Instead, one is constrained to accept the features selected by the statistical method (e.g. features classes in MAXENT; splits in tree-based methods) to represent that predictor (within some user-specified bounds). Rather, it is preferable to specify a set of features (or multiple sets for competing models) to determine the suitability for describing a particular pattern. For example, when features are selected automatically, it may be challenging to determine whether a quadratic term that makes the response unimodal is important or how much better/worse the model might be without it.

Complex

The starting premises, for hypothesis testing, is a priori ecological understanding enabling the user to select a small set of features. However, we do not always have this prior understanding. Complex models explore much larger sets of nonlinear features and interactions than simple models and are suited for generating hypotheses about underlying processes (Boulangeat et al. 2012) derived from potentially flexible responses that would not often be detected with simpler models (e.g. bimodality). This same flexibility can be used to augment existing knowledge. For example, if we know that a species is associated with dry, high elevation locations, we don't need a simplified model to describe this, but rather more insight from a potentially complex model to capture bimodality or strong asymmetries. Complex models also provide tools for evaluating predictor importance, which is useful for both generation and testing of hypotheses and can lead to inference that differs little from simpler 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).

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 or environmental space (cf. Peterson et al. 2011, Aarts et al. 2012). 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. The correlative models discussed here are not optimal for extrapolation in many cases; process-based models are generally preferred 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, Merow et al. 2014).

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 unintended ways and inadvertently model clustering in geographic space as complexity in environmental space, which can lead to errant interpolation (section Spatial autocorrelation).

Extrapolation always requires that response curves have been checked for biological plausibility (cf. section Niche description vs range mapping). Of course, even simple models can extrapolate poorly. For example, Thuiller et al. (2004) showed that a simple GLM or GAM run on a restricted and incomplete range could create spurious termination of the smoothed relationships, leading to errant extrapolation. Hence, the importance of extrapolation can depend on the chosen spatial extent and on the selected features (section Spatial extents and resolution). Complex models should be carefully monitored at the edges of the data range, both because small sample sizes and the ways different statistical methods handle extrapolation can have drastic effects on predictions (Pearson et al. 2006).

When using complex models, feature space may be sparsely sampled, which means that when one expects to interpolate a predictor, there may be inadvertent extrapolation of nonlinear features. For example, in a model with interaction terms, one may adequately sample the linear features for all predictors while poorly sampling the relevant combinations of these predictors (Zurell et al. 2012). Complex models can lead to different combinations of features producing similar model performance in the present (Maggini et al. 2006), but vastly diverging spatial predictions when transferred to other conditions (Thuiller 2003, Thuiller et al. 2004, Pearson et al. 2006, Edwards et al. 2006, Elith et al. 2010). Narrowing the range of possibilities using a simpler model that controls for the biological plausibility of the response curves (cf. section

Complexity in models) can reduce this divergence (Randin et al. 2006).

Data attributes

Sample size

The number of occurrence records is a critical limiting factor when building SDMs. With 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 approaches described earlier are potentially suitable, dependent on study objectives.

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 misspecified climate response surface. While simple models have a chance of smoothing over such variations, 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.

Sampling bias

Sampling bias arises from imperfect sampling design, which includes purposive, non-probabilistic, or targeted sampling (Schreuder et al. 2001, Edwards et al. 2006) and imperfect detection (MacKenzie et al. 2002). The important question is whether sampling bias – which often arises in geographic space – transfers to bias in environmental space, and further, whether some environments are completely unsampled. No statistical manipulation can fully overcome biased sampling. The main challenge when choosing complexity is that – particularly for models based on presence-only data – it may be unclear whether patterns in environmental space derive from habitat suitability, divergence between the fundamental and realized niches (Pulliam 2000), transient behavior, or sampling problems (Phillips et al. 2009, Hefley et al. 2013, Warton et al. 2013). For presence–absence data with perfect detection, sampling biases may not be too detrimental as long as at least some samples exist across environments into which the model is required to predict (Zadrozny 2004, but see Edwards et al. 2006 for contrasting results).

Recommendations

More flexible models will be more prone to finding patterns in restricted parts of environmental space where sampling is problematic. Poor performance on test data could identify over fitting to sampling bias, but only if the test data are unbiased. In practice, if unbiased testing data were available, they could be used to build an unbiased model in the first place. Recent advances that enable presence-only and presence–absence data to be modeled together, and across species, will be useful in this context (Fithian et al. 2014). A tradeoff exists between a complex model that might fit, e.g. step functions to few data points in poorly sampled regions and simple models that predict 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 (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).

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, the use of 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 (Austin and Nicholls 1997, Oksanen 1997, Austin 2002, 2007, Guisan and Thuiller 2005, Franklin 2010), which motivates constraining the functional form of response curves a priori (section Complexity in models; 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, Letten et al. 2013), and may be prone to overfitting with complex models if those data generally fail to describe the local habitat conditions accurately. One can use simple models to smooth over such idiosyncrasies if the main trends are sufficiently strong or one can 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. 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.

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 (Boulangeat et al. 2012, Dubuis 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 Sampling bias). The choice of extent can also determine whether the species entire range is included in the model or whether data are censored (e.g. limited by political borders).

Recommendations

Resolutions should be chosen that provide data from proximal rather than distal variables. Such data are becoming available at high resolutions with expanded and technologically enhanced monitoring networks and more sophisticated interpolation of climate data (e.g. PRISM). The choice of resolution hence reduces to the discussion of proximal versus distal predictors in section Predictor variables: proximal vs distal. When the extent is chosen to contain the species' entire range, models should include sufficient complexity to detect unimodal, skewed responses (section Complexity in models).

Simple

Smooth responses, characterized by simpler models, are to be expected at large spatial extents and coarse resolution that smooth over the confounding processes that affect finer resolution occurrence patterns (Austin 2007). At finer resolutions, it may also be undesirable to incorporate the full complexity of the response curve: much of the finer details may derive from factors for which no predictor variables are available or are irrelevant to the purpose of the investigation (e.g. microhabitat or regional competition effects).

Complex

At small spatial extents, we might have data on the relevant proximal factors (e.g. soil properties), so fitting complex models along small-scale gradients can capture this complexity. Also, complex models may be useful for exploring the nonlinearities that arise in response curves from distal variables at broad scales in that they potentially provide insight into important unmeasured variables.

Spatial autocorrelation

Many processes omitted from SDMs have spatial structure. For example, dispersal limitation, foraging behavior, competition, prevailing weather patterns, and even sampling bias can all lead to spatially structured occurrence patterns that are not explained by the set of 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, one should assess the degree of spatial autocorrelation in the residuals and implement methods to control for spatial autocorrelation. 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) cells, so it can effectively be ignored. One should be cautious building complex models because in practice, obtaining spatially independent cross-validation samples is extremely challenging when the underlying spatial process is unknown and failing to do so likely leads to over-fitting (cf. Hijmans 2012).

Complex

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

Conclusions

Methodological

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

Bias can come from over fitting complex models, and it can come from misspecified simple models. To find a model of optimal complexity, many approaches are possible and are readily justified if sufficient cross-validation has been performed. One might consider starting simple and adding the minimum complexity necessary (Snell et al. 2014, this issue), or conversely starting with a complex model and removing as much superfluous complexity as possible. If one can narrow down the potential complexity based on the considerations discussed here to consider models within a particular modeling approach (Table 1), then traditional model selection techniques are appropriate (section Modeling decisions).

Due to the exploratory nature of many SDMs and the desire to discover spatial patterns and their drivers, we recommend that analyses begin exploration using complex models to determine an upper bound on the complexity of response curves. Over fitting can be controlled through cross-validation (e.g. k-fold, and particularly block resampling methods), even if a full decomposition into train-validation-test data is not feasible. Furthermore, complex models can be used to identify smooth, simple occurrence–environment relationships if patterns are sufficiently strong and guide specification of simpler models. In contrast, it will be more difficult to overcome a misspecified simple model, should a more complex response exist. If the exploration with complex models reveals smooth relationships, one can shift to a simpler model. If instead strong nonlinearities are prevalent, one should consider biological explanations for the nonlinearities. If complex nonlinearities cannot be avoided, one should focus on minimizing the complexity, understanding it through sensitivity analysis and uncertainty analysis (below) and providing biologically based hypotheses about it. The end result is a model that adds complexity only to the extent necessary to reproduce observed patterns.

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

Biological

Despite the valuable insights we can gain from occurrence models, it is worth acknowledging that fundamental limitations to biological inference may emerge from these studies (Tyre et al. 2001, Araújo and Guisan 2006, Araujo and Peterson 2012, Merow et al. 2013). Balancing complex and simple models in such a way as to discover and discuss these limits may be as important as the actual patterns identified with some datasets. More broadly, it is important to keep in mind that we are ultimately performing exploratory analyses of occurrence–environment relationships. Occurrence records are not the ideal data to predict attributes of populations, Thuiller et al. (2014) provide an interesting cautionary note by showing weak relationships between occurrence probability and various demographic parameters for 108 tree species in temperate forests. However, 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 (Schurr et al. 2012, Thuiller et al. 2013, Snell et al. 2014).

A novel, and potentially important, application of SDMs is for informing mechanistic models about the shapes of response curve in demographic models (Merow et al. 2014), or dynamic spatio-temporal population models (Pagel and Schurr 2012, Boulangeat et al. 2014, Thuiller et al. 2014). 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. 2014). 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).

Many issues of response curve complexity that we discuss are also relevant for process-based SDMs. Representations of processes are incorporated into SDMs to improve the precision and accuracy, or to improve our understanding of ecological processes. Consequently, process-based models are used more for prediction and hypothesis testing than description and hypothesis generation. Yet, preferences for different model complexity persist (Evans et al. 2013, Lonergan et al. 2014). Study objectives influence the choice of complexity; i.e. whether the model is intended for extrapolation or for understanding the potential importance of mechanisms. In the former case, simple models are useful to make the study of the role of a mechanism more analytically tractable. In the latter case, preference might be towards more complex models, where the roles of specific mechanisms can be understood in relation to other interconnected mechanisms. When the objective is prediction, complex models are valuable to represent all known relevant mechanisms in order to obtain the 'best guess'. Simpler models are valuable when analyses imply that only certain key mechanisms are needed for sufficient predictive accuracy (further discussion in Evans et al. 2013). Attributes of the available data may be less important with process-based models when relevant test datasets are well understood. However, data considerations are important when mechanisms or parameters are inferred from data or when assessing the spatiotemporal resolution over which particular degrees of abstraction and parameter values are relevant (Evans et al. 2013, Lonergan 2014, Snell et al. 2014). In any case, we expect that progress towards improved process-based models lies in challenging occurrence-based SDMs with stronger biological justifications and interpretations that aim to shed light on the mechanisms that drive process-based models.

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 any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Aarts, G. et al. 2012. Comparative interpretation of count, presence–absence and point methods for species distribution models. – Methods Ecol. Evol. 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. – Funct. Ecol. 24: 1192–1201.

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

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

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

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

Austin, M. P. 1976. On non-linear species response models in ordination. – Vegetatio 33: 33–41.

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

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

Austin, M. P. and Smith, T. M. 1989. A new model for the continuum concept. – Vegetatio 83: 35–47.

Austin, M. P. and Nicholls, A. O. 1997. To fix or not to fix the species limits, that is the ecological question: response to Jari Oksanen. – J. Veg. Sci. 8: 743–748.

Austin, M. P. et al. 1994. Determining species response functions to an environmental gradient by means of a β-function. – J. Veg. Sci. 5: 215–228.

Barry, S. and Elith, J. 2006. Error and uncertainty in habitat models. – J. Appl. Ecol. 43: 413–423.

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

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

Boulangeat, I. et al. 2014. Anticipating the spatio-temporal response of plant diversity and vegetation structure to climate and land use change in a protected area. – Ecography 37: 1230–1239.

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

Breiman, L. et al. 1984. Classification and regression trees. – Wadsworth International Group.

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

Burnham, K. and Anderson, D. R. 2002. Model selection and multimodel inference: a practical information-theoretic approach. – Springer.

Chase, J. M. and Leibold, M. A. 2003. Ecological niches. – Univ. of Chicago Press.

Chave, J. 2013. The problem of pattern and scale in ecology: what have we learned in 20 years? – Ecol. Lett. 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. Biogeogr. 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. – J. Veg. Sci. 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. – Ecol. Model. 199: 132–141.

Elder, J. F., IV 2003. The generalization paradox of ensembles. – J. Comput. Graph. Stat. 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: 66–77.

Elith, J. and Leathwick, J. R. 2009. Species distribution models: ecological explanation and prediction across space and time. – Annu. Rev. Ecol. Evol. Syst. 40: 677–697.

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

Elith, J. et al. 2010. The art of modelling range-shifting species. – Methods Ecol. Evol. 1: 330–342.

Evans, M. R. et al. 2013. Do simple models lead to generality in ecology? – Trends Ecol. Evol. 28: 578–583.

Fithian, W. et al. 2014. A proportional observer bias model for multispecies distribution modeling. – arXiv in press.

Fox, J. 2003. Effect displays in R for generalised linear models. – J. Stat. Softw. 8: 1–27.

Franklin, J. 2010. Moving beyond static species distribution models in support of conservation biogeography. – Divers. Distrib. 16: 321–330.

Friedman, J. H. 2001. Greedy function approximation: a gradient boosting machine. – Ann. Stat. 29: 1189–1232.

Gelman, A. and Pardoe, I. 2007. Average predictive comparisons for models with nonlinearity, interactions, and variance components. – Sociol. Methodol. 37: 23–51.

Grömping, U. 2009. Variable importance assessment in regression: linear regression versus random forest. – Am. Stat. 63: 308–319.

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

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

Hastie, T. J. and Tibshirani, R. J. 1990. Generalized additive models. – Chapman Hall.

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

Hefley, T. J. et al. 2013. Nondetection sampling bias in marked presence-only data. – Ecol. Evol. 3: 5225–5236.

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. – Ecol. Model. 199: 142–152.

Holt, R. D. 2009. Bringing the Hutchinsonian niche into the 21st century: ecological and evolutionary perspectives. – Proc. Natl Acad. Sci. USA 106: 19659–19665.

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

Hutchinson, R. A. et al. 2011. Incorporating boosted regression trees into ecological latent variable models. – In: Burgard, W. and Roth, D. (eds), Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence. Association for the Advancement of Artificial Intelligence, pp. 1343–1348.

Ibáñez, I. et al. 2009. Multivariate forecasts of potential distributions of invasive plant species. – Ecol. Appl. 19: 359–375.

Janson, L. et al. 2013. Effective degrees of freedom: a flawed metaphor. – arXiv in press.

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

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

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

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

Lawson, C. R. et al. 2013. Prevalence, thresholds and the performance of presence–absence models. – Methods Ecol. Evol. 5: 54–64.

Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? – Ecology 74: 1659–1673.

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

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

Lonergan, M. 2014. Data availability constrains model complexity, generality, and utility: a response to Evans et al. – Trends. Ecol. Evol. 29: 301–302.

Lonergan, M. et al. 2014. Data availability constrains model complexity, generality, and utility: a response to Evans et al. – Trends. Ecol. Evol. 29: 301–302.

MacKenzie, D. I. and Royle, J. A. 2005. Designing occupancy studies: general advice and allocating survey effort. – J. Appl. Ecol. 42: 1105–1114.

MacKenzie, D. I. et al. 2002. Estimating site occupancy rates when detection probabilities are less than one. – Ecology 83: 2248–2255.

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

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

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

Merow, C. et al. 2014. On using integral projection models to build demographically driven species distribution models. – Ecography 37: 1167–1183.

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

Oksanen, J. 1997. Why the beta-function cannot be used to estimate skewness of species responses. – J. Veg. Sci. 8: 147–152.

Pagel, J. and Schurr, F. M. 2012. Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. – Global Ecol. Biogeogr. 21: 293–304.

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

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

Pearson, R. G. et al. 2013. Shifts in Arctic vegetation and associated feedbacks under climate change. – Nat. Clim. Change in press.

Peterson, A. T. et al. 2011. Ecological niches and geographic distributions. – Princeton Univ. Press.

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

Phillips, S. et al. 2006. Maximum entropy modeling of species geographic distributions. – Ecol. Model. 190: 231–259.

Phillips, S. et al. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. – Ecol. Appl. 19: 181–197.

Pulliam, H. R. 2000. On the relationship between niche and distribution. – Ecol. Lett. 3: 349–361.

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

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

Santika, T. and Hutchinson, M. F. 2009. The effect of species response form on species distribution model prediction and inference. – Ecol. Model. 220: 2365–2379.

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

Schurr, F. M. et al. 2012. How to understand species' niches and range dynamics: a demographic research agenda for biogeography. – J. Biogeogr. 39: 2146–2162.

Snell, R. et al. 2014. Using dynamic vegetation models to simulate plant range shifts. – Ecography 37: 1184–1197.

Soberón, J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. – Ecol. Lett. 10: 1115–1123.

Strobl, C. et al. 2008. Conditional variable importance for random forests. – BMC Bioinform. 9: 307.

Svenning, J.-C. et al. 2011. Applications of species distribution modeling to paleobiology. – Q. Sci. Rev. 30: 2930–2947.

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

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

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

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

Thuiller, W. et al. 2014. Does probability of occurrence relate to demographic performance? – Ecography 37: 1155–1166.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. – J. R. Stat. Soc. B 58: 267–288.

Tobalske, C. 2002. Effects of spatial scale on the predictive ability of habitat models for the green woodpecker in Switzerland. – In: Scott, J. M. et al. (eds), Predicting species occurrences: issues of accuracy and scale. Island Press, pp. 197–204.

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

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

Warton, D. I. et al. 2013. Model-based control of observer bias for the analysis of presence-only data in ecology. – PLoS One 8: e79168.

Welsh, A. H. et al. 2013. Fitting and interpreting occupancy models. – PLoS One 8: e52015.

Wood, S. N. 2006. Generalized additive models. – CRC Press.

Zadrozny, B. 2004. Learning and evaluating classifiers under sample selection bias. – ICML 2004: 114.

Zimmermann, N. E. et al. 2009. Climatic extremes improve predictions of spatial patterns of tree species. – Proc. Natl Acad. Sci. USA 106 (Suppl. 2): 19723–19728.

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

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