Title
Species distribution models in conservation biogeography: developments and challenges

Permalink
https://escholarship.org/uc/item/07p3949t

Journal
Diversity and Distributions, 19(10)

ISSN
1366-9516

Author
Franklin, Janet

Publication Date
2013-10-01

DOI
10.1111/ddi.12125

Peer reviewed
Species distribution models in conservation biogeography: developments and challenges

Janet Franklin

INTRODUCTION

Species distribution modelling (SDM) associates georeferenced observations of a biotic response variable – typically species occurrence or abundance – with multiple environmental predictors using a broad array of statistical learning methods (Elith & Leathwick, 2009; Franklin, 2010b; Elith & Franklin, 2013). A model estimated from observations can then be applied to digital maps of predictors resulting in a spatial prediction of the response variable, for example probability of species occurrence or habitat suitability. Species distribution modelling has deep roots in spatial decision support for land management (Hoffer, 1975; Kessell, 1976; Strahler, 1981), weed or pest species risk assessment (Sutherst & Maywald, 1985; Busby, 1991) and studies of climate impacts on the biota (Busby, 1986; Nix & Busby, 1986). On the order of a thousand papers a year are currently being published that use SDM methods, dramatically increased from about ten per year in the 1980s (Peterson & Soberon, 2012).

Species distribution modelling (sometimes called environmental or bioclimatic niche modelling) relies on ecological theory of processes that mediate species distributions and abundance – especially niche theory (Austin, 2002). In fact, burgeoning applications of SDM appears to have driven a renaissance in niche theory development and articulation (e.g. Godsoe, 2010). Species distribution modelling also relies on, and lies at the foundation of, three decades of development in geographic information science (GIScience) and remote sensing. It is obvious that SDM requires geospatial data for spatial prediction, but it has also driven developments in the field of GIScience. Environmental and terrain modelling has been identified as one of the three major subdomains in GIScience, with the most cited literature in that area including papers by M. F. Hutchinson, I. D. Moore, A. K. Skidmore, M. P. Austin and A. Guisan (see Figure 2 in Goodchild, 2010); this classic literature is directly related to species distribution modelling.

_Diversity and Distributions_ is a journal of conservation biogeography. Its mission is to publish papers that apply biogeographical principles, theories and methods (those addressing the distributional dynamics of taxa and assemblages) to problems concerning the conservation of biodiversity. The study of biological invasions is considered a key component of conservation biogeography, and the journal is an important forum for research on biogeographical aspects of biological invasions (Richardson, 2004; Richardson & Whittaker, 2010). _Diversity and Distributions_ has seen a steeply increasing trend in the number of submissions and published studies that use SDM as a method of analysis. Of the 579 papers published between January 2008 and June 2013 (queried 23 June 2013), for example, 100 (and an additional 16 accepted and posted on EarlyView) used SDM in some capacity – that is, they developed empirical models of species-environment correlations that were used to make a spatial prediction. Those papers addressed problems ranging from forecasting risk of future biological invasions, pathogen spread and climate change impacts, to spatial conservation planning and historical biogeography (Fig. 1). They span taxonomic groups, habitats and geographical regions (Appendix S1).

This editorial serves as an introduction to a virtual issue of _Diversity and Distributions_ that compiles key papers on species distribution modelling published in the journal (Table 1; Appendix S2). Papers selected for the virtual issue include contributions that both address pressing conceptual and methodological issues and provide key examples of the use of SDM for biodiversity assessment, conservation planning, risk analysis for invasive species and forecasting global climate change impacts. I selected those papers that have a high rate of citation relative to time since publication (empirical evidence that they are influential and useful; Table 1) or more recently published papers that are particularly creative in their use of SDM to support conservation biogeography (my subjective judgment or prediction that they will become influential). Another aim of the editorial is to suggest some profitable avenues of research relating to SDMs, both ‘nuts and bolts’ work on the philosophical underpinnings and technical aspects of such modelling, but also how SDMs could and should be used in advancing the aims of conservation biogeography.

The following sections describe the articles in the Virtual Issue and their linkages by grouping them into three areas: (1) those that address vexing methodological issues in SDM ranging from variability among modelling methods to sample size and sample design, (2) those that use SDM in innovative and rigorous ways to ‘interpolate’ species distributions in space, for example for biodiversity inventory, prospecting and conservation planning and (3) those that combine SDM with other data and methods in thoughtul ways to ‘extrapolate’ species distributions to different places or time periods to forecast impacts of environmental change on species...
I conclude with prospects and priorities for future research on modelling species distributions in support of conservation biogeography research.

**METHODODOLOGICAL ISSUES**

Methodological papers included in the virtual issue tended to focus on challenges that face modellers who must rely on pres-
ence-only observations of species occurrences, such as those available from natural history collections and increasingly from global databases that compile those collections information and other observations. This group of papers also addressed methodological issues of spatial dependence and non-stationarity, sample design, data resolution, sample size and consensus forecasting. Tsoar et al. (2007) compared six presence-only modelling methods, and while they did find systematic differences in performance among methods, also found that differences among species tended to be consistent across models. Osborne et al. (2007) showed that local regression methods are appropriate for interpolation of species distributions in space, while global methods are more appropriate for extrapolation to different places or time periods. Dark (2004) also demonstrated that spatial (auto-)regression models were more effective at identifying correlates of distribution of invasive species than non-spatial models.

Elith & Leathwick (2007) examined the effect of background sample design on model performance for presence-only models, finding that target group background performed better than a random sample, a conclusion borne out by subsequent studies. Guisan et al. (2007) found that a tenfold (single order of magnitude) change in spatial grain of data did not greatly affect SDM performance; there was only a slight trend towards lower performance at the coarser scale. Wisz et al. (2008) described how SDM performance degrades with smaller sample sizes. Consensus forecasting is one way of dealing with SDM uncertainty (Araújo & New, 2007). Marmion et al. (2009) evaluated five methods for calculating or deriving a consensus prediction from multiple SDMs and found that simple averaging, or accuracy-weighted averaging, of the probabilities estimated by different methods for the same data were the best-performing consensus methods.

Elith et al. (2011) provided an explanation of a widely used SDM algorithm for presence-only data (MaxEnt) in statistical terms and showed how the characteristics of species and species data affect model implementation decisions. They demonstrated that lack of absence data means that species prevalence cannot be estimated, sample selection bias has a strong effect on presence-only models (and there are ways to select background sample with same bias as presences), and that the way the extent of the region is defined (that the background sample is drawn from) also has big effect on these models.

Conceptual papers included in the Virtual Issue address the importance of establishing a strong conceptual framework for matching methods with data and questions (Jiménez-Valverde et al., 2008) and ways to use SDM in combination with other tools for forecasting or extrapolation (Franklin 2010a).

**INTERPOLATING SPECIES DISTRIBUTIONS IN SPACE USING SDM**

Species distribution modellings have proven to be powerful tools for conservation biogeography, especially when they are used for ‘interpolation’ – to fill in the geographical gaps in our knowledge of species distributions. This approach is effective when observations of species distributions are sparse, and correlations of those distributions with mapped environmental gradients are strong. Interpolation using SDM is very useful tool for biodiversity inventory, biodiversity prospecting (designing biodiversity surveys – predicting new occurrences), gap analysis, prioritizing areas for conservation (reserve design) and environmental impact analysis (determining how human activities including resource management might affect critical habitat for species of conservation concern).

Several recent examples of effective use of SDM to fill in the geographical gaps in species distributions are included in the virtual issue (Table 1). Platts et al. (2010) used SDMs to make spatial predictions of plant species richness for a biodiversity hotspot and suggested that because models are most uncertain for species of conservation concern, they should be developed iteratively with targeted fieldwork. Williams et al. (2009) took the next step in their study – several SDM methods were compared for their ability to predict the distributions of rare plant species, and further field surveys based on those predictions yielded discovery of new populations. Dubuis et al. (2011) compared direct statistical modelling of plant species richness as the response variable versus ‘stacking’ predictions from individual species models and concluded that both direct estimation (unbiased with correct response curve shape, but low accuracy) and the stacked approach (overestimating richness but yielding information about community composition) are complimentary and useful for conservation planning. Puschendorf et al. (2009) used SDM methods to predict the potential distribution of amphibian chytrid fungus in Costa Rica, a disease that threatens amphibians globally. Their study identified climatic (topographic) refuges where this pathogen, and therefore infectious outbreak, may be less likely; this information could be used for spatial conservation planning aimed at preserving amphibian diversity.

**EXTRAPOLATING SPECIES DISTRIBUTIONS ACROSS SPACE OR TIME**

Species distribution modellings are no longer enough on their own when we want to extrapolate, for example, the effects of future global (climate, land use) change on the biota, and risk of invasive species, although extrapolation has come to be their primary mode of application and has included studies remarkably broad in scope (Warren et al., 2013). Species distribution modellings are limited in their ability to forecast to novel environments by their empirical nature and equilibrium assumption, especially if naively applied with inadequate data. If used with explicit consideration of these limitations, however, they can be an important part of a methodological toolkit used to address pressing forecasting needs (Franklin, 2010a).
There are three ways SDMs can be more effectively used for extrapolation. (1) Data or information from more mechanistic or process-based studies or models (population, ecophysiology, community dynamics) can be incorporated during conceptual and statistical formulation (see Table 9.1 in Franklin, 2010b), for example deriving explanatory variables, variable selection, model estimation, specifying interactions and response curve shape (Elith et al., 2010). (2) SDMs can be linked with process models (Franklin, 2010a). This is sometimes called hybrid modelling (Dormann et al., 2012), but often the output from one model is used as the input to another, without feedback, so ‘linked’ or ‘coupled’ modelling is more descriptive. (3) Predictions from SDMs can be compared with process-based models and much can be learned from where and how they agree and disagree, in light of their respective assumptions (for example Kearney et al., 2010; Serra-Diaz et al., In Press).

In the virtual issue, Hof et al. (2012) used the first approach to extrapolation, informing their SDM with information about important biotic interactions affecting the distribution and abundance of the focal species (predator-prey dynamics). Naujokaitis-Lewis et al. (2013) used the second approach, linking SDM and population models to forecast climate change impacts in a study of Hooded Warbler to examine uncertainty due to different Global Climate Models (GCMs). Population viability estimates were sensitive to GCM effect on vital rates, but more sensitive to direct habitat loss projected from SDM. Thuiller et al. (2006) contrasted the predicted changes in plant species richness under climate change when assuming no dispersal versus unlimited dispersal from current distributions to bracket the range of outcomes likely to be generated using process models that more explicitly simulated dispersal.

The problem of extrapolation to predict risk of invasive species was addressed by Beaumont et al. (2009) who found that including data from the entire (native and non-native) distribution of invasive species may better characterize its fundamental niche and better forecast potential for invasion in space and time, for example, under climate change (but see Webber et al., 2011). Václavík & Meentemeyer (2012) used temporally explicit data on the invasion of a plant pathogen and demonstrated that SDMs calibrated at the early stages of invasion tend to underestimate the potential range compared with those calibrated with data from later stages. Dark (2004) sought to understand the environmental correlates of the distributions of invasive versus non-invasive non-native plant species in California (many of which are well established); she found the same factors to be important in both cases (lower elevations, higher road density and higher native plant species richness), pointing to the importance of species traits in determining whether an alien species is invasive or not.

Finally, a broad-reaching study recently published in Diversity and Distributions by Junker et al. (2012) modelled habitat suitability for African great ape taxa using environmental and human impact variables representing conditions in the 1990s. They projected these models to the 2000s based on updated human impact variables (population density, proximity to roads, etc.) and estimated losses of suitable habitat ranging from 11% to 59% for different taxa. While they cautioned that the coarse spatial scale of the analysis meant that it is informative to broad-, but not fine-scale conservation planning, their temporal extrapolation was short-term and well justified and was based on actual, observed changes in the driving variables (rather than modelled projections). This is an exemplary use of SDM for extrapolation over a limited time horizon in support of conservation biogeography.

**PROSPECTS AND PRIORITIES FOR THE FUTURE**

Because forecasting species distributions in novel or non-analogue environments is so central to conservation biogeography in an era of rapid global change (Sala et al., 2000), research that develops and tests innovative ways of forecasting impacts of global change – climate change, land use change, invasive species including emerging infectious diseases, altered disturbance regimes – on biodiversity should be of great interest to Diversity and Distributions. Hindcasting distributions to address historical and phylogeographical questions can also inform conservation biogeography (e.g. Porto et al., 2013; Smith et al., 2013). Molecular methods can provide genetic information about historical demography and dispersal dynamics of taxa (Scoble & Lowe, 2010; Duckett et al., 2013). Incorporating information about diversity below the species level may be particularly important for identifying genetically and geographically structured populations that may differ in their potential for genetic adaptation to environmental change (Hamann & Aitken, 2013) or for invading new regions (Thompson et al., 2011). In addition to genetically distinct populations, understanding factors driving the distributions of species’ functional types or traits (McGill et al., 2006; Kearney & Porter, 2009), as well as community properties such as taxonomic or phylogenetic diversity, effectively links ecological theory to conservation biogeography (e.g. Slik et al., 2009; Dubuis et al., 2011; Sypbard et al., 2013). Shifts in disturbance regimes that play out at large spatial scales may have important implications for conservation biogeography (Reside et al., 2012; Sypbard et al., 2013).

Moving forward, Diversity and Distributions is interested in publishing those studies that use insights from phylogeography and palaeodistribution dynamics, draw in cutting-edge work on genetics, and build on key developments in invasion, population and community ecology, to address critical information needs in conservation biogeography. These studies are likely to be multiscale and multidisciplinary, interfacing with climate and land change science, so that drivers of species distributions can be characterized at relevant scales. Species distribution modelling can be part of a methodological toolkit to address these information needs. Its limitations are well known, but solutions to those limitations are also
being described in the growing literature on this topic. Often overcoming those limitations involves collecting additional data about species, ecological communities and habitat (Elith & Franklin, 2013).

JANET FRANKLIN
School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ 85287-5302, USA
E-mail: janet.Franklin@asu.edu

REFERENCES


Hoffer, R.M. (1975) Natural resource mapping in mountainous terrain by computer analysis of ERTS-1 satellite data. Agricultural Experimental Station Research Bulletin 919, and LARS Contract Report 061575, Purdue University, W. Lafayette, IN.


and alien Australian acacias with correlative and mechanistic models. *Diversity and Distributions*, 17, 978–1000.


**SUPPORTING INFORMATION**

Additional Supporting Information may be found in the online version of this article:

**Appendix S1** Species distribution modelling (SDM) papers published in *Diversity and Distributions* Jan 2008 – May/Jun 2013.

**Appendix S2** Virtual Issue – Species Distribution Models in Conservation Biogeography: Developments and Challenges.