

Combining static and dynamic variables in species distribution models under climate change

Jessica C. Stanton¹, Richard G. Pearson^{2,3}, Ned Horning², Peter Ersts² and H. Reşit Akçakaya^{1*}

¹Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794-5245, USA; ²Center for Biodiversity and Conservation, American Museum of Natural History, New York City, NY, USA; and ³Department of Herpetology, American Museum of Natural History, Central Park West at 79th Street, New York, New York 10024, USA

Summary

1. Methods used to predict shifts in species' ranges because of climate change commonly involve species distribution (niche) modelling using climatic variables, future values of which are predicted for the next several decades by general circulation models. However, species' distributions also depend on factors other than climate, such as land cover, land use and soil type. Changes in some of these factors, such as soil type, occur over geologic time and are thus imperceptible over the time-scale of these types of projections. Other factors, such as land use and land cover, are expected to change over shorter timescales, but reliable projections are not available. Some important predictor variables, therefore, must be treated as unchanging, or static, whether because of the properties of the variable or out of necessity. The question of how best to combine dynamic variables predicted by climate models with static variables is not trivial and has been dealt with differently in studies to date. Alternative methods include using the static variables as masks, including them as independent explanatory variables in the model, or excluding them altogether.

2. Using a set of simulated species, we tested various methods for combining static variables with future climate scenarios. Our results showed that including static variables in the model with the dynamic variables performed better or no worse than either masking or excluding the static variables.

3. The difference in predictive ability was most pronounced when there is an interaction between the static and dynamic variables.

4. For variables such as land use, our results indicate that if such variables affect species distributions, including them in the model is better than excluding them, even though this may mean making the unrealistic assumption that the variable will not change in the future.

5. These results demonstrate the importance of including static and dynamic non-climate variables in addition to climate variables in species distribution models designed to predict future change in a species' habitat or distribution as a result of climate change.

Key-words: bioclimate, land use, MAXENT, niche model, SDM, soil

Introduction

Species distribution models (SDMs; or 'ecological niche models') were initially developed as single time step 'snapshots' of how species are distributed on the landscape. Increasingly, SDMs have been applied to assess the potential impacts of future climate change on biodiversity (e.g. Peterson *et al.* 2002; Thomas *et al.* 2004; Araújo, Thuiller & Pearson 2006) adding

an additional dimension of time to the approach. Adding this dimension requires that additional care be given to the selection of predictor variables and how those variables are used in these models.

Although several different algorithms have been applied (Elith *et al.* 2006), SDMs share a common generic approach (Hirzel *et al.* 2002): (i) the study area is divided into grid cells at a specified resolution; (ii) known species' presence localities (and sometimes absence localities) are used as the dependent variable; (iii) a number of environmental variables (e.g. temperature, precipitation, soil type, aspect, land cover type)

*Correspondence author. E-mail: akcakaya@life.bio.sunysb.edu
Correspondence site: <http://www.respond2articles.com/MEE/>

are gathered for each cell as predictor variables; and (iv) the suitability of each cell for the species is defined as a function of the environmental variables. The suitability of each cell can then also be estimated under changed environmental conditions, including scenarios of future climate change (Pearson & Dawson 2003). This essentially correlative approach is in contrast to more mechanistic models, which aim to directly model physiological relationships between climate variables and species responses. Both correlative and mechanistic methods have advantages and disadvantages, and it is generally acknowledged that both approaches have value in assessing the response of species to climate change (Kearney & Porter 2009; Buckley *et al.* 2010). For this study, we focus on the correlative approach.

Applications of correlative SDMs to estimate the impacts of climate change commonly characterize each grid cell using only climate variables (e.g. the 19 bioclimate variables available through the WorldClim dataset; Hijmans *et al.* 2005). These variables are then adjusted to reflect future climate scenarios (IPCC 2007), and in this sense, these variables are 'dynamic' because they change over the timeframe being modelled. SDMs built using only climate variables are commonly termed 'bioclimate envelope' models and are most commonly applied at large spatial scales (Pearson & Dawson 2003). Example applications of this climate-only approach include estimates of climate change impacts on amphibians and reptiles in Europe (Araújo, Thuiller & Pearson 2006; Carvalho *et al.* 2010), plants in Europe (Thuiller *et al.* 2005), birds in sub-Saharan Africa (Hole *et al.* 2009) and tropical rainforest vertebrates in Australia (Williams, Bolitho & Fox 2003).

However, species' occurrence is not only defined by climate variables, and exclusion of other important variables (such as soil type and land cover type) may reduce discriminatory ability, leading to inferior predictions (Iverson & Prasad 1998; Brook *et al.* 2009). Moreover, growing recognition of the importance of synergistic impacts between different threats to biodiversity emphasizes the need to assess risks posed by multiple factors acting together (Travis 2003; Brook, Sodhi & Bradshaw 2008). All of these different factors (or predictor variables) change through time, but the complication is that they may change at different rates or in unpredictable ways. For some variables, such as soil type, appreciable changes occur over geologic timescales and any changes over the time-scale being modelled (typically extending to the end of the 21st century for climate change impacts studies) are likely to be virtually undetectable. In other cases, we expect there will be changes in the variable, yet future scenarios are not available. This is particularly relevant for remotely sensed variables, such as land cover classifications and measures of productivity, which have proven important for estimating present-day species distributions (Zimmermann *et al.* 2007; Buermann *et al.* 2008). Predicting changes in land cover is difficult in part because land use patterns result from a confluence of factors including physical properties of the environment, resource demand, human population density and available technology, in addition to an array of laws, policies, mores and attitudes of people towards their physical environment. These factors are

ever-changing and can cause existing trends and patterns of land use to shift rapidly and in unexpected ways. In this study, we explore the best methods for using static and dynamic variables in species distribution modelling. Although it can be argued that no variable is truly static, for practical purposes, we define 'static' variables as those that are changing so slowly that cumulative change over the modelled time period is expected to be negligible (such as soil), as well as those that may be changing at a faster rate but for which future projections are either not available or not reliable (such as land use), although current, reliable maps are available. We define 'dynamic' variables (e.g. bioclimatic variables) as those that are expected to change substantially over the modelled time period, and for which reliable, or at least generally accepted future projections are available, even if those projections are uncertain.

Approaches for combining dynamic and static variables in SDM predictions for future climate projections remain poorly understood and contentious (Brook *et al.* 2009). While some authors have included only climate variables (e.g. Williams, Bolitho & Fox 2003; Thuiller *et al.* 2005; Araújo, Thuiller & Pearson 2006; Hole *et al.* 2009; Carvalho *et al.* 2010), others have included non-climatic, static predictor variables. For example, Peterson *et al.* (2002) included elevation, slope and aspect alongside temperature and precipitation variables when predicting the impacts of future climate change on Mexican faunas. In another example, when modelling trees in the eastern United States, Iverson & Prasad (1998) tested models built using (i) only climate variables, and (ii) climate variables alongside edaphic, land use/land cover and elevation variables. Iverson and Prasad concluded that the best models included a mixture of climatic and non-climatic variables, and they, therefore, predicted distributions under dynamic climate by including both dynamic and static variables in the models.

Differing opinions as to whether and how static and dynamic variables should be combined have been driven, in part, by alternative views as to the role of non-climatic variables in correlative models. One view stresses that only including climate variables could cause the model to be overly sensitive to climate change under future climate scenarios (Iverson & Prasad 1998). The alternative view states that including non-climatic, static variables could result in models that are well-fit to current distributions yet will be insensitive to future climate scenarios because climate variables are down weighted in these models. The situation is complicated by various possible interactions between static and dynamic variables, in particular, the problem of correlations between predictors. To take an extreme example, consider the effect of including elevation alongside temperature as environmental variables in SDMs. Elevation *per se* does not have a direct physiological effect on species, but rather tends to be strongly correlated with factors that may have direct physiological effects such as air pressure, temperature and precipitation. Thus, adjusting temperature to reflect a future climate scenario (i.e. temperature is dynamic) while keeping elevation static will cause significant

inconsistencies in a SDM, due because of changes in the correlation structure between these two variables (Austin 2002). Inclusion of an indirect and static variable (elevation) that is strongly correlated with, and a proxy for, a direct and dynamic variable (temperature) is clearly problematic in this instance, although the situation may be less clear when correlations are not so obvious (for instance, between climate and land cover; Thuiller, Araújo & Lavorel 2004) or when the variable has a direct physiological effect on the species (such as solar radiation derived from a digital elevation model (DEM); Austin & Van Niel 2011).

Species occurrences may also depend on how static and dynamic variables interact. For instance, a hypothetical plant species may occur on all soil types when precipitation (or water availability) is above a certain value, but only on a subset of soil types when the precipitation is below this value. Applying SDMs in situations where suitability is dependent on both static and dynamic variables, and there are likely to be correlations and dependencies between them, raises important methodological questions: Should a mixture of static and dynamic environmental variables be included in the model? Should variables that are expected to change in the future be included even if future scenarios are not available? Might predictions be improved by modelling only with dynamic variables and then using static variables to mask out areas that are unsuitable because of non-climatic factors?

Here, we explore these issues using the *MAXENT* SDM approach (Phillips, Anderson & Schapire 2006), which provides a powerful method for fitting complex species–environment relationships, can incorporate interactions between different variables and shows good predictive performance when compared to alternative SDM approaches (Elith *et al.* 2006). All SDM methods have limitations concerning, for example, extrapolating to environments not included in model calibration (Pearson *et al.* 2006) and dealing with spatial bias in occurrence records (Graham *et al.* 2008), yet such issues have been relatively well explored for *MAXENT* (e.g. Phillips & Dudík 2008; Phillips *et al.* 2009). To test model performance, we used simulated (artificial) species whose environmental requirements can be defined so as to test model performance precisely (*sensu* Elith & Graham 2009). Our goal is to provide practical methodological guidance on the concurrent use of static and dynamic variables in SDMs and thereby to contribute toward the development of general standards on this issue within the SDM community.

Methods

To evaluate different approaches to handling both dynamic and static variables in a predictive SDM model as realistically as possible, we used simulated species to replicate how a real species would be modelled in a comparable situation. For example, when assessing the effect of dynamic land use, we assumed the modeller could not know how the landscape would change and thus can either include land use as a static layer or exclude it from the model (see details later). However, we assumed that the modeller would know which climatic variables (precipitation, temperature) and which static variables (soil,

vegetation cover) are important components of a species' habitat. We did this because our goal is to assess different ways of combining static and dynamic variables, and we did not want to compound this with uncertainty resulting from lack of knowledge about the variables contributing to suitability. Similarly, we assumed that there is no error in predictions of future change in climatic variables, because we did not want to compound the results with uncertainty about climate projections. In a real application, uncertainty about the appropriate habitat variables would be incorporated by developing alternative models with different sets of predictive variables. Similarly, uncertainty about climate projections would be incorporated by using alternative climate change scenarios.

DATA SOURCES AND ENVIRONMENTAL LAYERS

Although we constructed simulated species for this study, we used real environmental variables on an actual landscape to define the niche space. We used the North American continent, roughly west of the Mississippi river (*c.* 90°W) to the Pacific coast (*c.* 125°W), for the landscape. Four climate variables (maximum temperature of the warmest month, minimum temperature of the coldest month, precipitation of the driest month and precipitation of the wettest month) were extracted from the WorldClim database at 30 arc-seconds (*c.* 1 km by 1 km) resolution (Hijmans *et al.* 2005). Future climate scenarios for the 2050s and 2080s, based on an A2a emissions scenario and the HadCM3 climate model, were available on the WorldClim website (<http://www.worldclim.org>).

To simulate the effects of changes in land use for the true habitat suitability, we used human population density, crop and pasture layers from the HYDE History Database of the Global Environment (Klein Goldewijk, Beusen & Janssen 2010; Klein Goldewijk *et al.* 2011). The HYDE project uses historical records of human populations to estimate the demand for agricultural land based on the available technology of the time period. Land for crops and pasture is then allocated spatially, based on the estimated demand and suitability of the landscape for those uses. The HYDE layers for population density, cropland and pasture are mapped globally at a 5 min-by-5 min (*c.* 9.5 km × 9.5 km) grid resolution. Population is mapped as average density per square kilometre. The crop land and pasture layers are each mapped as the number of square kilometres in each respective land use per grid cell.

We included a variable for soil type based on the Harmonized World Soil Database (Fischer *et al.* 2008). We reclassified the original categories within the study area based on similarity of water retention capacity and fertility characteristics. We categorized four soil types: porous, low-porosity humic, sand/gravel and saturated soils. We also identified glaciers and salt flats.

To demarcate areas in the landscape that are relatively less disturbed by human modification, we created a vegetation cover layer as a static, categorical variable. This variable is based on vegetation classification classes from the North America Land Cover Characteristics Database, Version 2.0, (NALCC2), which is part of a global land cover database derived from Advanced Very High Resolution Radiometer (AVHRR) satellite data (Loveland *et al.* 2000). Vegetation type and land cover classes from the NALCC2 database were collapsed to identify areas of the landscape that are largely natural vegetation considered suitable for our simulated species. Developed urban areas and land primarily used for agriculture were considered unsuitable.

We geographically aligned and scaled the resolution of the three HYDE layers, soil type and vegetation cover to match the 30 arc-second resolution of the WorldClim layers.

SIMULATED SPECIES

We created three different simulated species to evaluate the performance of alternative modelling approaches in cases when the predictor variables interact with each other or change through time. When mapping the future suitability, we assumed niche conservatism under climate change. Therefore, for each species, we created future suitability maps with the same suitability function as described later for all time steps, using the projected future climate variables corresponding to each time step mapped.

To determine the best approach for handling a static variable that interacts with dynamic variables in the determination of habitat suitability, we created a species named INTERACTING, for which habitat suitability was partially determined by an interaction between soil type (which influences water holding capacity) and precipitation variables (Fig. 1). In contrast to INTERACTING, we created the second species NON-INTERACTING whose habitat suitability is determined by the same predictor variables as INTERACTING, but where soil type and precipitation are independent factors in determining habitat suitability (Fig. 2). For each species simulated, we adjusted the functions so that the suitable area for the species was not too small, too large or outside the modelled region at any of the time periods. Maps of habitat suitability for both simulated species are available in Supporting Information Fig. S1.

Land cover and land use variables can be important for determining species distributions, and including them in models can improve model performance beyond the use of climate variables alone (Pearson, Dawson & Liu 2004). To address the best way to combine land use variables with projections of future climate, we constructed a third species called SENSITIVE that is sensitive to human modification of the landscape, i.e., it avoids areas of high population density and crop and pasture land. To model this species' 'true' suitability through time, we paired the projected climate variables with the HYDE land use layers for North America between the dates of 1930

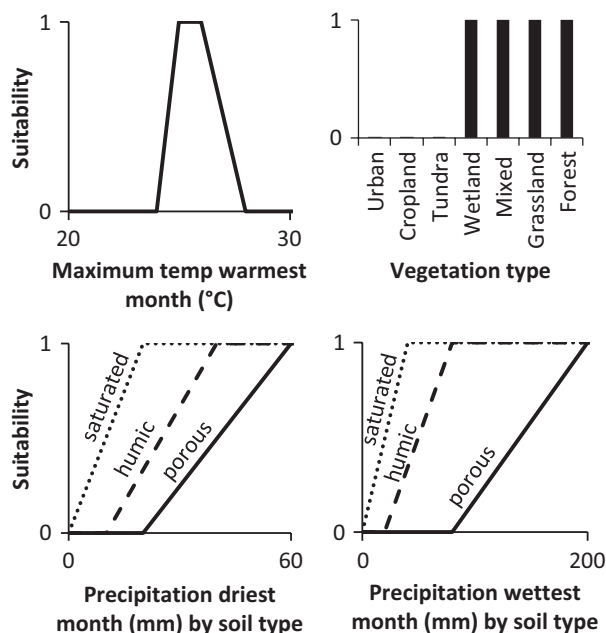


Fig. 1. Habitat suitability function defining response of the INTERACTING organism to each environmental variable. Total suitability at a given location is the product of each of the individual functions illustrated.

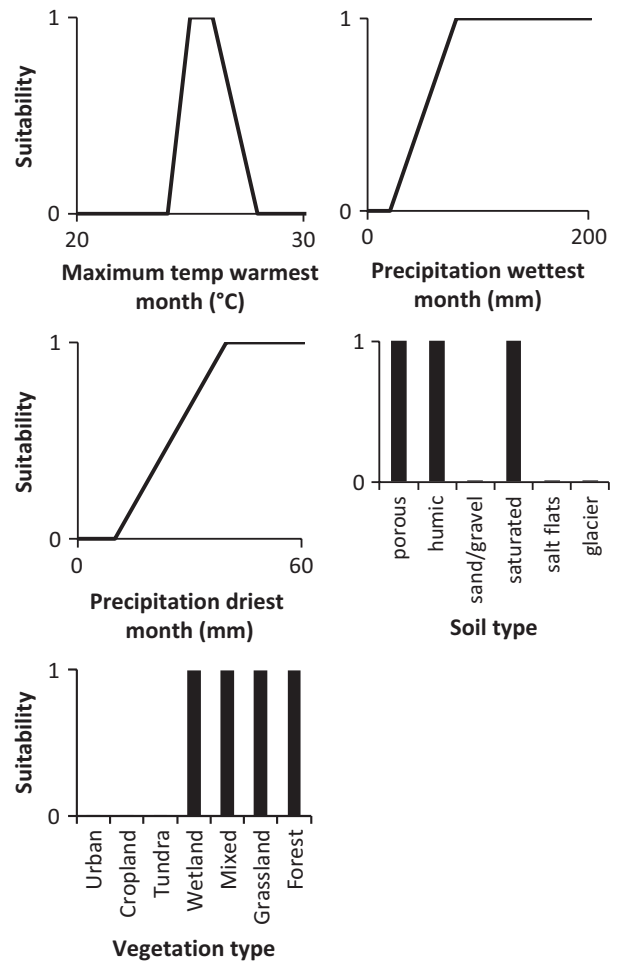


Fig. 2. Habitat suitability function defining response of the NON-INTERACTING organism to each environmental variable. Total suitability at a given location is the product of each of the individual functions illustrated.

and 2000. We chose to use an historical trajectory of dynamic land use because it is an actual representation of a relationship between a society and its physical landscape. By directly using the observed land use through time, we can include a level of complexity and nuance that would likely be lacking from a land use model simply fitted to current trends and projected into the future. We selected the time period beginning with 1930 because by this time in North America's history, most of the population centres and major agricultural production areas had already been established and it was not our intention to model an initial human pioneering event. This organism's true habitat suitability function is described in Fig. 3 (for mapped habitat suitability, see Fig. S2). Note that the true future suitability for the SENSITIVE species was constructed using the time steps for 1930, 1970 and 2000, but when the organism was modelled, the land use variables were treated as static using only land use variables for 1930 for each projected future time step. This was performed to demonstrate the treatment of a variable which is likely to change in the future, but in unknowable ways, and as a result can either be modelled as a static variable or excluded from the model.

SPECIES DISTRIBUTION MODEL CONSTRUCTION

For each species, we sampled *c.* 200 occurrence locations from the mapped true habitat suitability for 2010. We selected this number of

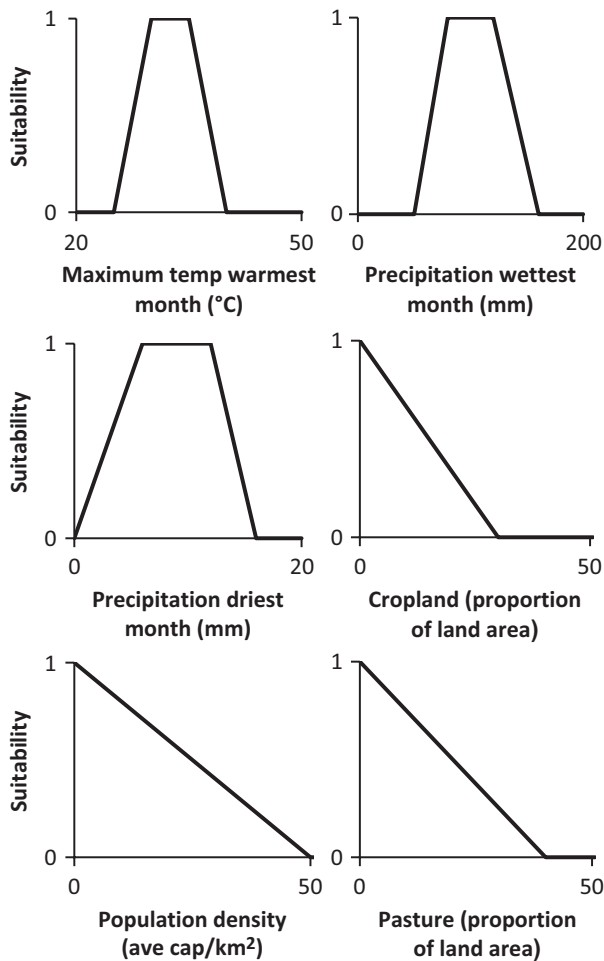


Fig. 3. Habitat suitability function defining response of the SENSITIVE organism to each environmental variable. Total suitability at a given location is the product of each of the individual functions illustrated.

occurrences as an attempt to balance the need to have enough locations to adequately fit the habitat suitability function with the reality that in practice occurrence locations for many species are not available in great numbers for SDM models. The occurrence locations were sampled randomly such that each grid cell on the map had a probability of being selected as an occurrence location proportional to the true habitat suitability value at that location. We then constructed a model for habitat suitability using MAXENT (ver. 3.3.2, Phillips, Anderson & Schapire 2006) treating the static variables (soil type, land cover or land use) as described later. MAXENT estimates suitability of the landscape for a species by fitting a function to the given occurrence locations and the predictor variables. Finally, we used the function from MAXENT fitted to the current climatic conditions to create predicted habitat maps for the future time periods of 2050 and 2080. Specifically, this was performed by applying the fitted functions to future climate variables using the 'projection' capability in the MAXENT software.

Each static variable was handled in one of the three ways in the modelled habitat suitability; (i) included directly in MAXENT as a predictor variable; (ii) excluded as a predictor variable from the model fitted by MAXENT, but with non-suitable areas masked from the final mapped habitat suitability; or (iii) excluded entirely from the model, meaning

the model was based solely on dynamic variables. We created mask layers for the static variables by characterizing each layer as a Boolean image of either suitable or non-suitable areas of the landscape. For the categorical variables (soil type and land cover; Figs 1 and 2), each category was simply considered suitable or not based on the true suitability definition for each species. For the HYDE land use variables (population density, crop area and pasture area; Fig. 3), which are continuous, it was necessary to assign a threshold value defining the boundary between what is considered suitable and non-suitable. We defined two levels of threshold values for each of the HYDE variables, one which was more restrictive and the other which was more liberal with regard to the total area masked from the modelled habitat suitability. We set threshold values for each land use variable at the point which corresponds to a true suitability value of 0.5 (for the restrictive land use mask) or 0.0 (for the liberal land use mask). After applying the threshold to each land use variable, we multiplied the three resulting masks to create a single Boolean land use mask where each cell is either suitable or non-suitable. When applying any of the masks, the continuous modelled habitat suitability values in locations overlapping with suitable areas in the mask layer were unchanged, while modelled suitability values overlapping with non-suitable masked areas were set to zero.

We repeated the occurrence location sampling and habitat suitability modelling steps for each of the three species 100 times so that the model results would not be dependent on any single random sample of the true habitat suitability map. Results were summarized over the 100 iterations for each of the three simulated species.

We evaluated the three different approaches for handling the static variables in the models by comparing the modelled present and future habitat suitability maps to the true habitat suitability maps for the corresponding time step. We evaluated the performance of different models using the area under the receiver operator characteristic curve (AUC) as a measure of model discrimination ability and the correlation coefficient as a measure of calibration (Pearce & Ferrier 2000). To calculate AUC, which compares the model fitted suitability values to presence/absence data, we randomly sampled 10 000 presence and 10 000 absence locations from the true suitability map for each species, where the probability of being selected as a presence location was proportional to the suitability value of that location. The correlation coefficient was calculated over all terrestrial pixels of the study area.

Results

Including static variables in the model with the dynamic variables for all three simulated species performed better or no worse than either masking or excluding the static variables (Figs 4 and 5).

The static variable that was interacting with the dynamic variables in the true habitat suitability function (soil type for INTERACTING model) showed more pronounced differences between treatments than non-interacting static variables (Fig. 4). When the static variable (soil type) was interacting with the climate variables, masking it rather than including it directly in the model reduced the fit by an average of 33%, as measured by the correlation coefficient. In contrast, when soil type was not interacting with climatic variables, there was a slightly improved fit (<1%) when masked as opposed to included in the model as a variable. Excluding the static variable (soil type) altogether reduced the fit more when it was

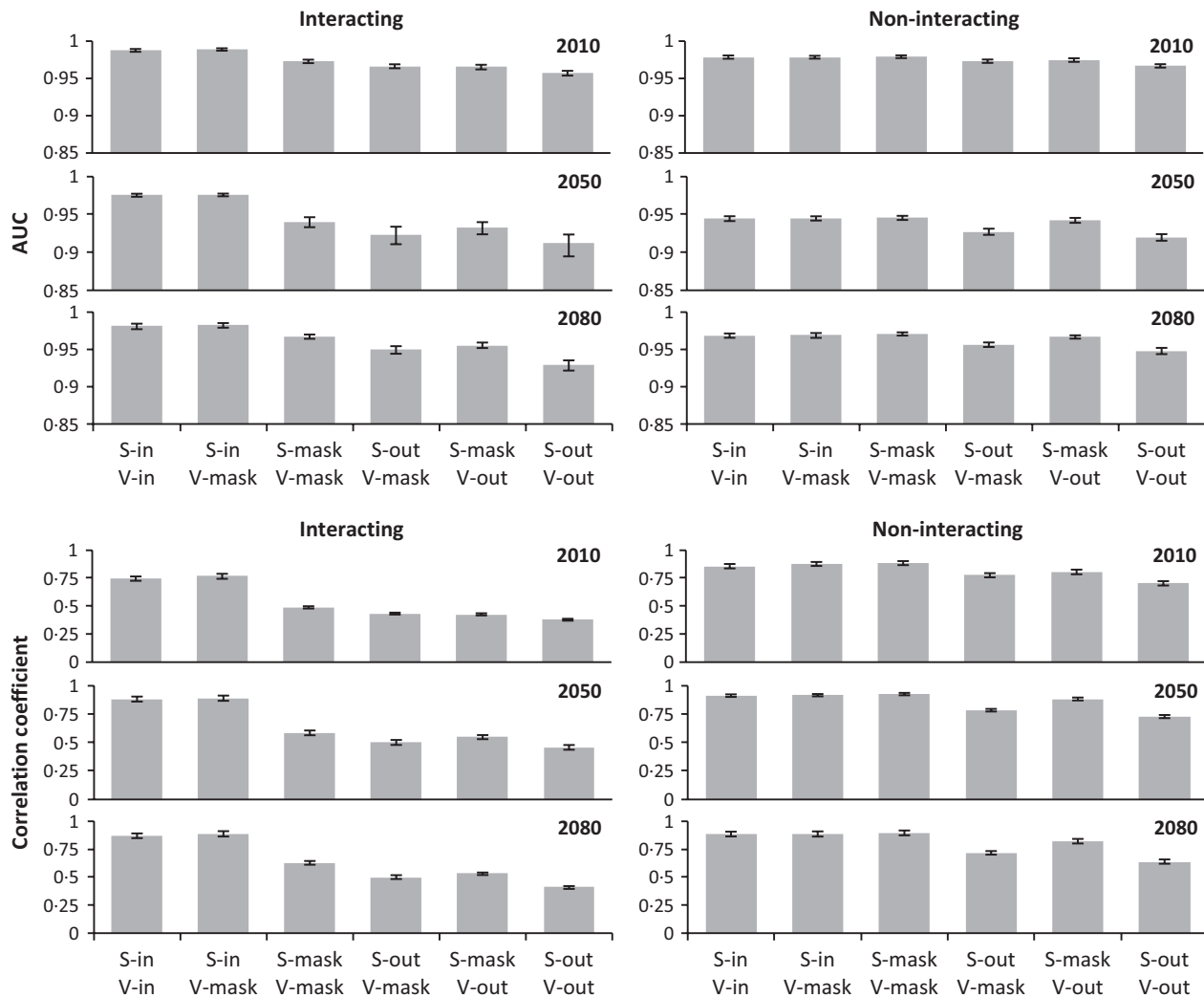


Fig. 4. Area under the receiver operator characteristic curve and correlation coefficients for the various treatments of the static variables soil type (S), and vegetation cover (V), where soil type is either interacting or non-interacting with climate variables in the true habitat suitability function. Static variables were: (i) included in the species distribution model (SDM) along with dynamic variables (in); (ii) used to mask unsuitable areas from the results of SDM built including only the dynamic climate variables (mask); or (iii) excluded entirely from the model (out). Results are summarized over the 100 simulations of each treatment. Bars illustrate the mean; error bars show the 5th and 95th percentiles.

interacting with the climatic variables: correlation coefficient was 44% lower when the interacting static variable was excluded, but only 15% when the non-interacting variable was excluded. A similar pattern is also observed for AUC, but with less variability between treatments.

Including vegetation cover resulted in model performance that was similar to masking the variable for both the INTERACTING and NON-INTERACTING simulated species. Excluding the vegetation cover from the model reduced the fit, as measured by the correlation coefficient, by an average of 13% over the three time steps vs. masking the variable when the soil type is excluded for INTERACTING and 9% for NON-INTERACTING.

Including static land use in the model for the simulated species SENSITIVE resulted in better model performance than either masking or excluding land use from the model (Fig. 5). Masking land use with the more liberal mask reduced the fit compared with including it as a model variable as measured by

the correlation coefficient by an average of 16% over the three time steps. The more restrictive mask reduced the fit by an average of 18%, and excluding land use altogether reduced the fit by 33%, averaged over all time steps.

Discussion

The results of this study demonstrate the importance of including static and dynamic non-climate variables in addition to climate variables in SDMs designed to predict future change in a species' habitat or distribution as a result of climate change. It is especially important to include variables that may interact with climate variables directly in the model. Using such variables (e.g. soil types) as a mask would make the invalid assumption that their effects on species distributions are independent of climate variables, whereas including them in the analysis allows the statistical SDM approach used (e.g. MAXENT) to incorporate

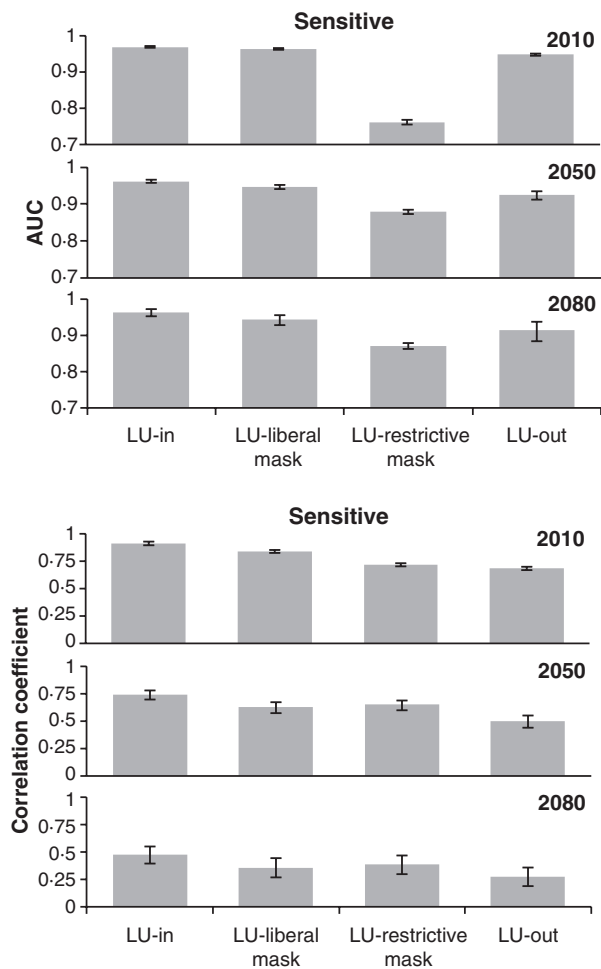


Fig. 5. Area under the receiver operator characteristic curve and correlation coefficients for the various treatments of land use (population density, cropland and pasture land) intensity as a static variable. The three land use variables were: (i) included in the species distribution model along with dynamic variables (in); (ii) used to mask unsuitable areas with either a loose or restrictive definition of unsuitable (liberal/restrictive mask; see text for details); or (iii) excluded entirely from the model (out). Results are summarized over the 100 simulations of each treatment. Bars illustrate the mean; error bars show the 5th and 95th percentiles.

interactions (i.e. dependencies) between them and the climate variables.

Some SDM methods (such as MAXENT) include interaction terms automatically or by default, whereas for other methods (e.g. GLM), interactions may have to be specifically added, for example, by creating and adding variables that are the product of two variables. Although we only used one modelling approach, we believe our general conclusions apply to other methods as long as they are applied to include interactions.

A fourth option that we did not test is to create a separate suitability layer based on the static variable (e.g. by assigning a separate suitability value to each soil type or running a separate SDM with the static variables only) and multiplying this map with the probability map that is output from the SDM with the climatic variables. This is similar to masking, albeit with more than two values for the 'mask' layer, and the multiplication

assumes that the two layers are independent or non-interacting; thus, we believe the implications for bioclimatic modelling are the same as masking.

A related approach, which has been used to integrate data at different spatial scales (Pearson, Dawson & Liu 2004), is to combine dynamic climate and static land cover data in a two-step process: (i) a climate-only model is built and shifts under future climate scenarios are predicted; (ii) the output from the climate-only model is used alongside land cover as inputs to a second model. This approach uses both static and dynamic variables in the same model, which is supported by our results here, and may be a useful way to integrate the large scale effects of climate with the more local effects of land cover (Pearson, Dawson & Liu 2004).

A complicated decision when projecting models to future climates involves how best to handle variables, such as human land use, that are expected to change in the future, but that are difficult or impossible to predict for future years. Clearly, using only the current data layers of such variables, in combination with dynamic climate layers, does not fully account for their effect on the future habitat suitability of the species. In such cases, it may be argued that leaving such variables out of the analysis might be better. However, our results indicate that if such variables do affect species distributions, including them in the model is better, even if it means making the unrealistic assumption that their values will not change in the future. We found the inclusion of static variables in the model improved performance for the present distribution and resulted in no or only small degradation in the predictive performance for future distributions. This suggests that some 'down-weighting' of climate variables may be appropriate for species with non-climatic influences on their habitat to avoid overestimating the effects of climate change. This was also true, to a lesser extent, when non-climate variables are included as masks, as long as the mask is not very restrictive. The overall poorer performance of the restrictive mask (i.e. one that excludes all but a small area as habitat) is probably because masking is in general a cruder and thus more error-prone form of including a variable. An interesting result of our analysis is that the more restrictive mask based on human land use resulted not only in overall poorer correlation with the true habitat suitability, but also in relatively better predictions in the future compared with the present. This is possibly because as human land use intensified through time, it eventually came to resemble the landscape of the static restrictive mask more than in the present.

In addition, static variables that only indirectly influence species' distributions and are highly correlated with climate variables should be excluded from the analysis. Static variables such as elevation, latitude or longitude may serve as useful proxies for current climatic conditions but can hinder the accuracy of future predictions as the relationships between the static and dynamic variables change in the future. Although not specifically tested in this study, a consideration of first principles reveals that including variables such as elevation directly in the SDM is likely to result in models in which the projected effects of future climate change are underestimated. In these instances, it is preferable to carefully consider each

environmental predictor variable and include those that are justifiably believed to be directly biologically relevant to the species. However, it should be noted that a number of possibly useful variables can be derived from a DEM as terrain can greatly influence factors such as temperature, orographic lifting, solar radiation, hydrology and air pressure (Moore, Grayson & Ladson 1991). Note that although correlation between elevation and temperature is a problem within SDMs, it is actually a benefit when modelling (or, interpolating) temperature and precipitation variables based on data from weather stations. Thus, elevation is used in the WorldClim data set as an independent variable for modelling temperature and precipitation (Hijmans *et al.* 2005).

In summary, for studies designed to predict future change in a species' habitat or distribution as a result of climate change, we recommend:

- 1 Static variables that are highly correlated with climate variables, and which have only indirect influences on species distributions, such as elevation, be excluded;
- 2 Static variables that are known or suspected to interact with climate variables, such as soil, be included in the analysis as additional explanatory variables (i.e. as input layers);
- 3 Static variables that are not expected to interact with climate variables be either included in the model as additional variables or used as a mask to remove areas that are not suitable;
- 4 Dynamic non-climate variables (e.g. those related to human land use) that are expected to change in the future be either included in the model as additional variables or used as a mask to remove areas that are not suitable, even if future change in these variables cannot be predicted, and thus, only the current maps can be used (if these variables are used as a mask, we recommend that the mask is not overly restrictive).

A promising recent application of SDMs is to link them with population dynamic models so as to estimate extinction risk (Keith *et al.* 2008; Anderson *et al.* 2009). Such risk estimates are dependent on realistic projections of species habitats, which are functions of not only the climatic variables that are commonly used in such models but also of many other variables (such as soil type, land use, vegetation cover, etc.), which are becoming increasingly available through remote sensing (Horning *et al.* 2010). Our analyses demonstrate the importance of incorporating these variables in SDMs and show that the best way to combine such variables with climate variables is to include them as explanatory variables.

Acknowledgements

This material is based upon work supported by the National Aeronautics and Space Administration under Grant No. NNX09AK19G issued through the NASA Biodiversity Program.

References

Anderson, B., Akçakaya, H.R., Araújo, M., Fordham, D., Martinez-Meyer, E., Thuiller, W. & Brook, B.W. (2009) Dynamics of range margins for metapopulations under climate change. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 1415–1420.

Araújo, M.B., Thuiller, W. & Pearson, R.G. (2006) Climate warming and the decline of amphibians and reptiles in Europe. *Journal of Biogeography*, **33**, 1712–1728.

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

Austin, M.P. & Van Niel, K.P. (2011) Impact of landscape predictors on climate change modelling of species distributions: a case study with *Eucalyptus fastigata* in southern New South Wales, Australia. *Journal of Biogeography*, **38**, 9–19.

Brook, B., Sodhi, N. & Bradshaw, C. (2008) Synergies among extinction drivers under global change. *Trends in Ecology & Evolution*, **23**, 453–460.

Brook, B.W., Akçakaya, H.R., Keith, D.A., Mace, G.M., Pearson, R.G. & Araújo, M.B. (2009) Integrating bioclimate with population models to improve forecasts of species extinctions under climate change. *Biology Letters*, **5**, 723–725.

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

Buermann, W., Saatchi, S., Smith, T.B., Zutta, B.R., Chaves, J.A., Milá, B. & Graham, C.H. (2008) Predicting species distributions across the Amazonian and Andean regions using remote sensing data. *Journal of Biogeography*, **35**, 1160–1176.

Carvalho, S.B., Brito, J.C., Crespo, E.J. & Possingham, H.P. (2010) From climate change predictions to actions – conserving vulnerable animal groups in hotspots at a regional scale. *Global Change Biology*, **16**, 3257–3270.

Elith, J. & 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., Graham, C., Anderson, R., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R., Huettmann, F., Leathwick, J., Lehmann, A., Li, J., Lohmann, L., Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J., Peterson, A., Phillips, S., Richardson, K., Scachetti-Pereira, R., Schapire, R., Soberon, J., Williams, S., Wisz, M. & Zimmermann, N. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Fischer, G., Nachtergaele, F., Prieler, S., Velthuisen, H.T., Verelst, L. & Wiberg, D. (2008) *Global Agro-ecological Zones Assessment for Agriculture (GAEZ 2008)*. IIASA, Laxenburg, Austria and FAO, Rome, Italy.

Graham, C.H., Elith, J., Hijmans, R.J., Guisan, A., Townsend Peterson, A. & Loiselle, B.A. (2008) The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology*, **45**, 239–247.

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

Hirzel, A.H., Hausser, J., Chessel, D. & Perrin, N. (2002) Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology*, **83**, 2027–2036.

Hole, D.G., Willis, S.G., Pain, D.J., Fishpool, L.D., Butchart, S.H.M., Collingham, Y.C., Rahbek, C. & Huntley, B. (2009) Projected impacts of climate change on a continent-wide protected area network. *Ecology Letters*, **12**, 420–431.

Horning, N., Robinson, J.A., Sterling, E.J., Turner, W. & Spector, S. (2010) *Remote Sensing for Ecology and Conservation: A Handbook of Techniques*. Oxford University Press, USA.

Intergovernmental Panel on Climate Change (IPCC) (2007) *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Edited by S. Solomon *et al.* Cambridge Univ. Press, Cambridge, UK.

Iverson, L.R. & Prasad, A.M. (1998) Predicting abundance of 80 tree species following climate change in the eastern United States. *Ecological Monographs*, **68**, 465–485.

Kearney, M. & Porter, W. (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters*, **12**, 334–350.

Keith, D.A., Akçakaya, H.R., Thuiller, W., Midgley, G.F., Pearson, R.G., Phillips, S.J., Regan, H.M., Araújo, M.B. & Rebelo, T.G. (2008) Predicting extinction risks under climate change: coupling stochastic population models with dynamic bioclimatic habitat models. *Biology Letters*, **4**, 560–563.

Klein Goldewijk, K., Beusen, A. & Janssen, P. (2010) Long-term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1. *The Holocene*, **20**, 565–573.

Klein Goldewijk, K., Beusen, A., van Drecht, G. & de Vos, M. (2011) The HYDE 3.1 spatially explicit database of human induced global land use

- change over the past 12,000 years. *Global Ecology and Biogeography*, **20**, 73–86.
- Loveland, T.R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, J., Yang, L. & Merchant, J.W. (2000) Development of a global land cover characteristics database and IGBP DISCover from 1-km AVHRR data. *International Journal of Remote Sensing*, **21**, 1303–1330.
- Moore, I.D., Grayson, R.B. & Ladson, A.R. (1991) Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. *Hydrological Processes*, **5**, 3–30.
- Pearce, J. & Ferrier, S. (2000) Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, **133**, 225–245.
- Pearson, R.G. & Dawson, T.P. (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*, **12**, 361–371.
- Pearson, R.G., Dawson, T.P. & Liu, C. (2004) Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography*, **27**, 285–298.
- Pearson, R.G., Thuiller, W., Araújo, M.B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T.P. & Lees, D.C. (2006) Model based uncertainty in species range prediction. *Journal of Biogeography*, **33**, 1704–1711.
- Peterson, A.T., Ortega-Huerta, M.A., Bartley, J., Sanchez-Cordero, V., Soberon, J., Buddemeier, R.H. & Stockwell, D.R.B. (2002) Future projections for Mexican faunas under global climate change scenarios. *Nature*, **416**, 626–629.
- Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.
- Phillips, S.J. & Dudík, M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161–175.
- Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J. & Ferrier, S. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, **19**, 181–197.
- Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F.N., de Siqueira, M.F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A.S., Midgley, G.F., Miles, L., Ortega-Huerta, M.A., Townsend Peterson, A., Phillips, O.L. & Williams, S.E. (2004) Extinction risk from climate change. *Nature*, **427**, 145–148.
- Thuiller, W., Araújo, M.B. & Lavorel, S. (2004) Do we need land-cover data to model species distributions in Europe? *Journal of Biogeography*, **31**, 353–361.
- Thuiller, W., Lavorel, S., Araújo, M.B., Sykes, M.T. & Prentice, I.C. (2005) Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 8245–8250.
- Travis, J.M.J. (2003) Climate change and habitat destruction: a deadly anthropogenic cocktail. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 467–473.
- Williams, S.E., Bolitho, E.E. & Fox, S. (2003) Climate change in Australian tropical rainforests: an impending environmental catastrophe. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **270**, 1887–1892.
- Zimmermann, N.E., Edwards, T.C., Moisen, G.G., Frescino, T.S. & Blackard, J.A. (2007) Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. *Journal of Applied Ecology*, **44**, 1057–1067.

Received 6 June 2011; accepted 17 August 2011

Handling Editor: Robert Freckleton

Supporting Information

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

Fig. S1. Habitat suitability maps for the simulated species INTERACTING and NON-INTERACTING, for 2010, 2050, and 2080. Habitat suitability ranges from 0 (unsuitable) to 1 (optimal).

Fig. S2. Habitat suitability maps for the simulated species SENSITIVE, for 2010, 2050, and 2080. Habitat suitability ranges from 0 (unsuitable) to 1 (optimal).

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be reorganized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.