

The projection of species distribution models and the problem of non-analog climate

Matthew C. Fitzpatrick · William W. Hargrove

Received: 25 August 2008 / Accepted: 20 January 2009 / Published online: 14 April 2009
© US Government 2009

By 2100, a quarter or more of the Earth's land surface may experience climatic conditions that have no modern analog, with novel climates predicted to arise primarily in regions that currently support high levels of biodiversity (Williams et al. 2007). Further, global commerce will continue to transport species beyond long-standing dispersal barriers, potentially unleashing biological invaders into regions outside of those in which they evolved. Global climatic change and biological invasions will each have important and likely synergistic impacts on biodiversity. However, the emergence of non-analog climates (i.e., climatic conditions that do not presently exist) and the introduction of species to new biogeographical settings challenge our ability to anticipate these impacts because little information exists to predict how species may respond under novel environments.

This problem is particularly relevant for projections in space and time made from species distribution models, which increasingly are being applied to conservation issues related to biodiversity and global change. Species distribution models use relationships developed between the observed distribution of the species and corresponding environmental conditions to predict the potential distribution of the species (for a recent review see Guisan and Thuiller 2005). Once developed from the current distribution, the model can be extrapolated in space to anticipate biological invasions (e.g., Peterson et al. 2004), or time to forecast potential changes in distribution of species under climatic change (e.g., Fitzpatrick et al. 2008), or both to forecast the potential for invasion under climatic change (e.g., Roura-Pascual et al. 2004).

M. C. Fitzpatrick (✉)
Harvard University, Harvard Forest, 324 North Main Street, Petersham, MA 01366, USA
e-mail: mfitzpat@fas.harvard.edu

M. C. Fitzpatrick
Department of Biological Sciences, University of Rhode Island, 100 Flagg Road, Kingston,
RI 02881-0816, USA

W. W. Hargrove
Eastern Forest Environmental Threat Assessment Center, USDA Forest Service,
Southern Research Station, Asheville, NC 28804-3454, USA

The validity of such forecasts is subject to many widely acknowledged uncertainties (Pearson and Dawson 2003; Thuiller 2004; Guisan and Thuiller 2005; Araújo and Rahbek 2006; Heikkinen et al. 2006; Pearson et al. 2006; Williamson 2006; Fitzpatrick et al. 2007), but one factor that has received less attention is the extrapolation of models into environments unlike those characterizing the region in which the model was calibrated (but see Thuiller et al. 2004). Because climatic shifts may create ‘new’ environments comprised of combinations of conditions that did not previously occur, especially when combined with local biogeographical and edaphic settings (Hargrove and Hoffman 2004; Saxon et al. 2005), this “problem” is potentially common in projections of species distribution models.

Forecasting future distributions of species from current species-climate relationships is problematic because the observed distribution of a species alone provides no information about how the species might respond under novel environments. Making a prediction under such novel conditions is not only prone to error (Heikkinen et al. 2006; Williamson 2006) it is also ecologically and statistically invalid. Although this issue is a recognized problem in the literature, relatively few studies have addressed it directly (but see Saetersdal et al. 1998; Ficetola et al. 2007).

There are multiple approaches to determine and visualize non-analog conditions (e.g., Williams et al. 2007). Here we propose a simple method using a modification of techniques already employed to project species distributions across space and time, which can be readily implemented by anyone familiar with such techniques. Although our method is amenable to most any statistical approach, it is particularly applicable to algorithms that are relatively opaque or ‘black box’ in character and which provide minimal insight to the fitted relationships on which spatial projections are based.

Figure 1 shows a Venn diagram representing a simplification of the multivariate environmental spaces encountered when projecting species distribution models. The large black circle on the left labeled ‘I’ represents the current combination of environmental conditions upon which the model is calibrated. This calibration region most appropriately represents environments within a biome or region to which the modeled species is endemic. The dashed circle defines the subset of these conditions within this biome or region under which the species has been observed. The large gray circle on the right labeled ‘II’ represents the expected future combination of environmental conditions to which the model is projected. This projection region presents either a potential host range in which the risk of invasion is assessed or the calibration region under a future climate scenario to assess potential impacts from climate change.

The goal of projecting species distribution models is to discriminate region A (the dotted portion of circle ‘II’ representing climatic conditions in the future range where the species is likely to be present) from region B (the hashed portion of circle ‘II’ representing climatic conditions in the future range where the species is likely to be absent). Models often report a binary prediction of presence and absence (often derived from continuous output). However, a third possibility is “no prediction possible” or a “null prediction.” A null prediction should occur in any area where the model must extrapolate to novel environmental conditions that have no analog to those combinations under which the model was calibrated. Such conditions are shown as the gray region of circle II labeled ‘C’ in Fig. 1.

It is general practice to not determine or report areas representing non-analog environments (region C, Fig. 1). Instead studies typically extrapolate models into non-analog conditions and assume such extrapolations are valid (but see Saetersdal et al. 1998; Ficetola et al. 2007). Some algorithms, notably maximum entropy (Maxent, Phillips et al. 2006), automatically deal with this issue by constraining the upper and lower bounds of future values of environmental variables to the range under which the model was calibrated, an

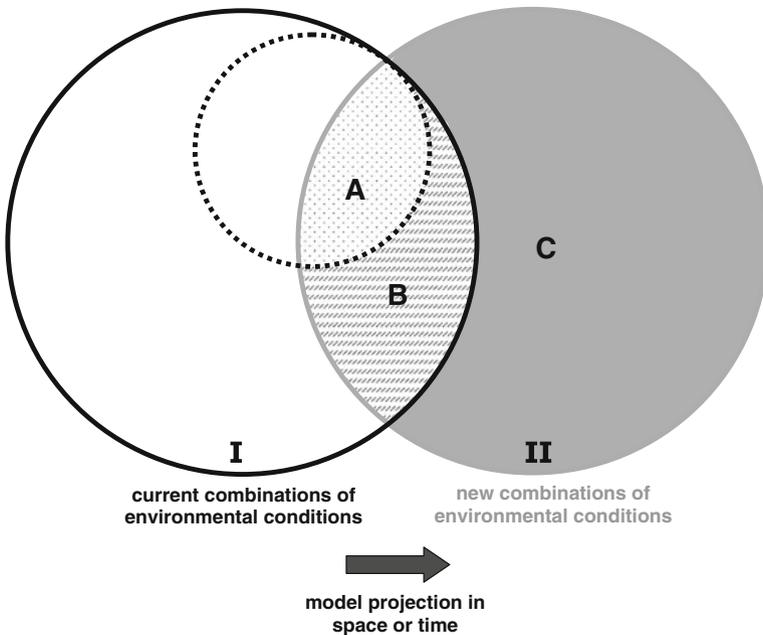


Fig. 1 Two-dimensional delineation of multivariate environmental spaces encountered when projecting species distribution models to forecast biological invasions and range shifts under climate change. The *large circle* on the *left* labeled 'I' represents environmental conditions upon which the model is calibrated (i.e., a biome or other region to which the modeled species is endemic), while the *dashed circle* represents the subset of conditions in which the species is present. The *large gray circle* on the *right* labeled 'II' represents environments to which the model is to be projected, either as a potential host range or a future climate scenario. The overlap of these *three circles* delineates three combinations of environmental conditions relevant to the projection of species distribution models: *A* suitable in the future range (predicted present), *B* unsuitable in the future range (predicted absent) and *C* non-analog (no prediction possible or null prediction)

approach Phillips et al. (2006) termed 'clamping'. However, sequential univariate clamping may not identify multivariate *combinations* of non-analog future conditions.

Failure to identify regions having non-analog environments can result in misinterpretation of potential future distributions of species (Thuiller et al. 2004). In effect, models may predict the species to be absent in areas that are otherwise suitable or may identify regions as highly suitable simply due to inappropriate extrapolation of response curves. These implied errors are not conservative, and each has important consequences for management of biological invasions and climate change impacts on biodiversity. In short, projections of species loss under climate change may be overestimated and/or areas to set aside for future conservation may be misidentified. In the context of biological invasions, regions at risk may be underestimated and/or areas predicted to be highly at risk simply may represent statistical artifacts. The conceptual problem of non-analog climate is not specific to any single algorithm, but is widespread across most species distribution modeling methods.

We suggest that a simple approach to determine and visualize areas where no prediction should be made (region C, Fig. 1) is to calibrate a model on the entire study region. In other words, consider all locations within the large black circle in Fig. 1 as presences and the remaining locations within the study area as absences and then project this model to the future environment. This method will model all combinations of current environmental

conditions found within the current range of the species and, when projected, will identify the overlap of these conditions with future environments in environmental space. When such a model is mapped in geographic space, regions containing non-analog environments are revealed (i.e., as areas of predicted absence), allowing such areas to be readily reported in conjunction with projected distributions. When the distribution of the study species itself is modeled and projected, its future distribution can be most reliably predicted within the zone of overlap between current and future environments.

Predictions should not be attempted at locations outside the projected distribution of the study region because these areas have environmental conditions that differ from conditions found within the environmental space in which the species-level model will be calibrated. Comparison of the projected range of the study species to that of the projected study region serves to determine and visualize non-analog environments. We call such a companion analysis to the projection of a species distribution model a “power of prediction analysis,” since it indicates the limits of the predictive ability of the resulting range projection in a spatially explicit way. Like a continuous projection of the distribution of the species itself, which can be converted to presence/absence using an appropriate threshold, the resulting projection from a power of prediction analysis is also continuous and can also be converted to a binary map delineating regions where models can be projected (prediction possible) from regions where projection should not be attempted (no prediction possible). Alternatively, continuous output can be interpreted as an indicator of the confidence of the ability to predict presence/absence of the modeled species. In fact, continuous output is arguably more desirable since some degree of extrapolation may be possible and a hard cut-off may not be appropriate. However, it is difficult to elucidate just how much extrapolation, if any, is warranted without detailed study. We argue it is better to at least indicate where extrapolation has occurred rather than report a spurious projection.

The extent of the region on which the model is calibrated should contain the complete gradient of environmental space that the study species could reasonably encounter, including consideration of dispersal ability and major biogeographical barriers or transitions. We acknowledge that delineating this area may not be obvious in many instances. For the sake of simplicity and to demonstrate the approach, we present two cases where the total possible extent of the distribution of the modeled species is relatively easy to define: an inland water body (the Caspian Sea), which has exchanged numerous species with the Great Lakes, and the southwestern Australia, a global biodiversity hotspot bounded by a steep precipitation gradient.

We projected a model of the entire Caspian Sea onto the Great Lakes using BIOMOD (Thuiller 2003; Thuiller et al. 2009) in R version 2.7.2 (R Development Core Team 2008). Within BIOMOD, six statistical techniques were employed to develop an ensemble forecast (Araújo and New 2007), including artificial neural networks (ANN), classification trees (CTA), generalized additive models (GAM), generalized linear models (GLM), mixture discriminate analysis (MDA), and random forest (RF). We selected ~7,500 equally spaced points across the entire Caspian Sea and an equal number of absence points from a background encompassing the eastern Atlantic and the Mediterranean, Baltic and Black Seas (Fig. 2a, inset). We used six environmental variables derived from satellite remote sensing at 4-km resolution. Mean, minimum and maximum annual temperature were derived from the Advanced Very High Resolution Radiometer (AVHRR) using monthly climate records collected during the period between 1985 and 2002. Data collected by the Moderate Resolution Imaging Spectroradiometer (MODIS) during 2001–2005 were used to construct the remaining three variables describing relevant physical characteristics of aquatic ecosystems: chlorophyll *a* concentration (a measure of productivity), diffuse attenuation coefficient, and

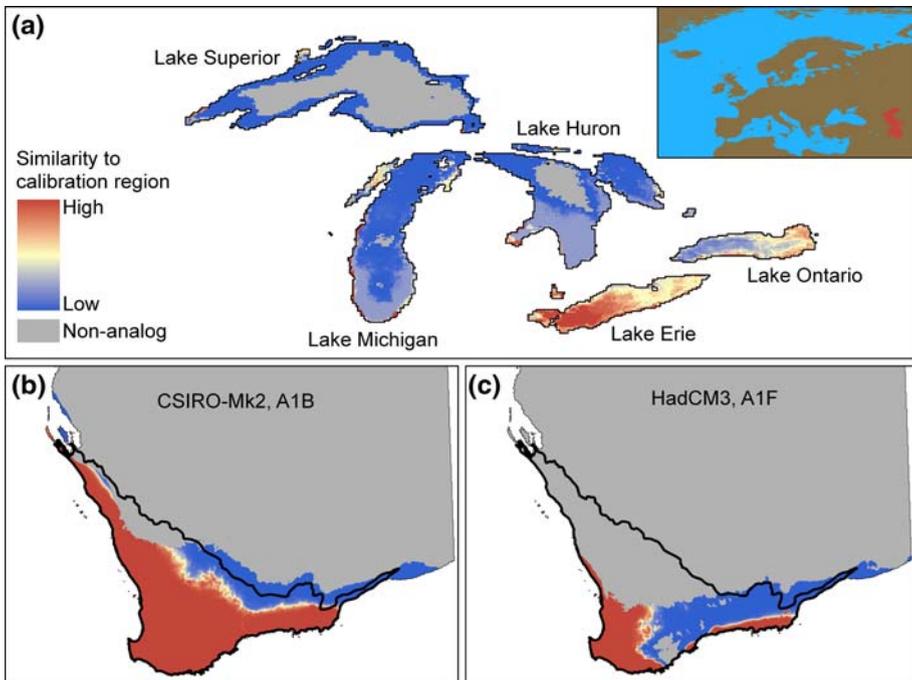


Fig. 2 Two demonstrations of analyses to determine non-analog environments for **a** a potential aquatic invasive species from the Caspian Sea (shown in red, inset) to the Great Lakes and **b, c** for climate change impacts under two future climate scenarios for the southwest biome of Western Australia. Warmer colors indicate conditions with high similarity to the calibration region and areas where models may reliably be projected, while blue shading indicates regions of low similarity and areas where models may be less reliably projected. Gray shading indicates non-analog environments and regions where predictions should not be made (i.e., locations where the modeled probability = 0)

normalized water-leaving radiance, (measures of water turbidity). For this example, the species-level model would be calibrated only on conditions within the Caspian Sea for a species endemic to the Caspian Sea.

Results of this analysis suggest that it may not be possible to predict confidently the vulnerability of areas in the interiors of Lakes Superior, Huron, and Michigan to aquatic invasive species from the Caspian Sea, as close analogs to such environments do not exist anywhere within these areas. The interiors of Lakes Michigan and Huron show limited similarity to any environments in the Caspian Sea (blue shading, Fig. 2a), whereas the interior of Lake Superior and portions of Lake Huron are completely non-analogous to the Caspian Sea in terms of this set of environmental variables (gray shading, Fig. 2a). In contrast, near shore areas of several Great Lakes and most of Lake Erie show relatively high degrees of similarity to the Caspian Sea (warmer colors, Fig. 2a). The ability to predict presence/absence generally decreases as distance from shore increases, suggesting that mainly near-shore environments in the Great Lakes are similar to environments found in the Caspian Sea, at least in terms of the variables used here. Distributional predictions within these near-shore areas are most reliable based on models developed from the Caspian Sea.

We performed a similar analysis for a biome and global biodiversity hotspot within Western Australia (bold line Fig. 2b, c) under climate change using seven climate variables

at 2.5-km resolution: mean annual temperature, minimum temperature of the coldest month, maximum temperature of the warmest month, annual, winter (June, July, August) and summer (December, January, February) precipitation, and an index of growing season length. We projected the biome onto two future climate scenarios for 2080, including the CSIRO-Mk2 model scaled using the IPCC A1B emission scenario and the HadCM3 model scaled using the IPCC A1F emission scenario. See Fitzpatrick et al. (2008) for details regarding environmental datasets and development of future climate scenarios. We calibrated the model using all cells within the biome as presences and all remaining cells within Western Australia as absences. For this example, the species-level model would be calibrated entirely within the boundaries of the hotspot for a species endemic to this area.

Our results suggest that by 2080 either none, or most, of the hotspot will disappear, depending on climate scenario, and will not re-appear elsewhere in Western Australia. Under the A1B scenario (Fig. 2b), potential impacts on biodiversity may be predicted reliably (within the limits imposed by other uncertainties in the modeling process) across most of the biome except along the northwestern and southeastern border. In contrast, presence/absence may not be reliably predicted across most of the biome under the A1F scenario (Fig. 2c).

A second interpretation is that some areas in southwestern Australia that become “non-analogous” in the future actually represent a southwestwardly expansion of existing conditions within the adjacent central arid Eremean biome (Fitzpatrick et al. 2008). In this sense, the issue is not with the development of non-analog climates per se, but rather with restricting the spatial domain upon which the species-level model is calibrated. This problem could be alleviated by increasing the size of the calibration region to ensure that the model is not used to extrapolate outside the calibration data range (Pearson et al. 2002). However, this approach will not work if the future conditions are novel globally (*cf.* Williams et al. 2007) or may become computationally prohibitive if the calibration region must cover a large spatial domain in order to capture the full range of future environmental conditions.

Non-analog environments may be prevalent across both space and time. We do not intend to suggest that it is impossible to forecast into non-analog conditions under some circumstances. Indeed, based on other information it may be possible to rule out non-analog climates as uninhabitable, but this may be next to impossible for most species using present knowledge. Rather, we argue that it is best practice to indicate the limitations of the model by determining and presenting areas where reliable projections cannot be made. Otherwise in reporting projections of species distribution models without consideration of non-analog climate conditions, ecologists may be misrepresenting the potential impacts of climate change and the geographic extent of biological invasions. Just as means should be reported with their corresponding confidence intervals, we suggest that projections from species distribution models should be paired with matching power of prediction analyses. Given the growing reliance on species distribution models to provide forecasts of the potential impacts of global climatic change and biological invasions on biodiversity, we argue the problems presented by non-analog environments to such forecasts warrants increased attention.

Acknowledgments MCF acknowledges support from the University of Tennessee in the form of a Yates Dissertation Fellowship and through the Department of Ecology and Evolutionary Biology. WWH thanks the Climate Simulation Group within the Computer Science and Mathematics Division at Oak Ridge National Laboratory for his guest status there. Don Catanzaro provided the aquatic environmental layers used in the Caspian Sea analysis. Comments from Rob Dunn, Rebecca Efroymson, Jack Williams, and three anonymous reviewers improved an early draft of this manuscript. The Australia Research Council (via an ARC grant to JD Majer and RR Dunn) and the US Environmental Protection Agency (via TN & Associates) supported portions of this research.

References

- Araújo MB, New M (2007) Ensemble forecasting of species distributions. *Trends Ecol Evol* 22:42–47
- Araújo MB, Rahbek C (2006) How does climate change affect biodiversity? *Science* 313:1396–1397
- Ficetola GF, Thuiller W, Miaud C (2007) Prediction and validation of the potential global distribution of a problematic alien invasive species: the American bullfrog. *Divers Distrib* 13:476–485
- Fitzpatrick MC, Weltzin JF, Sanders NJ et al (2007) The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? *Glob Ecol Biogeogr* 16:24–33
- Fitzpatrick MC, Gove AD, Sanders NJ et al (2008) Climate change, plant migration, and range collapse in a global biodiversity hotspot: the *Banksia* (Proteaceae) of Western Australia. *Glob Chang Biol* 14:1–16
- Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. *Ecol Lett* 8:993–1009
- Hargrove WW, Hoffman FM (2004) The potential of multivariate quantitative methods for delineation and visualization of ecoregions. *Environ Manag* 34:S39–S60
- Heikkinen RK, Luoto M, Araújo MB et al (2006) Methods and uncertainties in bioclimatic envelope modeling under climate change. *Prog Phys Geogr* 30:751–777
- Pearson RG, Dawson TP (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Glob Ecol Biogeogr* 12:361–371
- Pearson RG, Dawson TP, Berry PW et al (2002) Species: a spatial evaluation of climate impact on the envelope of species. *Ecol Model* 154:289–300
- Pearson RG, Thuiller W, Araujo MB et al (2006) Model-based uncertainty in species range prediction. *J Biogeogr* 33:1704–1711
- Peterson AT, Scachetti-Pereira R, Hargrove WW (2004) Potential geographic distribution of *Anoplophora glabripennis* (Coleoptera: Cerambycidae) in North America. *Am Midl Nat* 151:170–178
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecol Model* 190:231–259
- R Development Core Team (2008) R foundation for statistical computing, Vienna, Austria
- Roura-Pascual N, Suarez AV, Gomez C et al (2004) Geographical potential of Argentine ants (*Linepithema humile* Mayr) in the face of global climate change. *Proc R Soc Lond B Biol Sci* 271:2527–2534
- Saetersdal M, Birks HJB, Peglar SM (1998) Predicting changes in Fennoscandian vascular-plant species richness as a result of future climatic change. *J Biogeogr* 25:111–122
- Saxon E, Baker B, Hargrove WW et al (2005) Mapping environments at risk under different global change scenarios. *Ecol Lett* 8:53–60
- Thuiller W (2003) BIOMOD—optimizing predictions of species distributions and projecting potential future shifts under global change. *Glob Change Biol* 9:1353–1362
- Thuiller W (2004) Patterns and uncertainties of species' range shifts under climate change. *Glob Chang Biol* 10:2020–2027
- Thuiller W, Brotons L, Araújo MB et al (2004) Effects of restricting environmental range of data to project current and future species distributions. *Ecography* 27:165–172
- Thuiller W, Lafourcade B, Engler R et al (2009) BIOMOD—a platform for ensemble forecasting of species distributions. *Ecography* 32:1–5
- Williams JW, Jackson ST, Kutzbach JE (2007) Projected distributions of novel and disappearing climates by 2100 AD. *Proc Natl Acad Sci USA* 104:5738–5742
- Williamson M (2006) Explaining and predicting the success of invading species at different stages of invasion. *Biol Invasions* 8:1561–1568