Title
CONSEQUENCES OF SPATIAL AND TEMPORAL CLIMATE VARIABILITY FOR SPECIES DISTRIBUTION MODELING

Author
Fernandez, Miguel

Publication Date
2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA MERCED

CONSEQUENCES OF SPATIAL AND TEMPORAL CLIMATE VARIABILITY FOR SPECIES DISTRIBUTION MODELING

A dissertation submitted in partial satisfaction of the requirements
for the degree Doctor in Philosophy

in

Environmental Systems

by

Miguel Alejandro Fernandez Trigoso

Committee in charge:

Dr. Lara M. Kueppers, Chair
Dr. Michael N Dawson
Dr. Healy Hamilton
Dr. Anthony L. Westerling

2013

The Dissertation of Miguel Alejandro Fernandez Trigoso is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

Dr. Michael N Dawson

_____

Dr. Healy Hamilton

_____

Dr. Anthony L. Westerling

_____

Dr. Lara M. Kueppers, Chair

University of California, Merced
2013

When I heard the learn'd astronomer;

When the proofs, the figures, were ranged in columns before me;

When I was shown the charts and diagrams, to add, divide, and measure them;

When I sitting heard the astronomer where he lectured with much applause in the lecture-room,

How soon, unaccountable, I became tired and sick;

Till rising and gliding out, I wander'd off by myself,

In the mystical moist night-air, and from time to time,

Look'd up in perfect silence at the stars.

Walt Whitman, 1865

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CURRICULUM VITAE

My career is deeply rooted in the emerging science of conservation biogeography. I focus on the synergy between climate science, systems theory, geographic information science, biodiversity and natural resource management. I use computers to integrate geospatial information, ecological niche theory and the principles of evolution to characterize the uncertainty associated to ecological forecasting. My goal is to produce science-based management strategies for adaptation to human induced global change.

Education:

2003 Bachelor of Sciences, Biology concentration in Ecology, Universidad Mayor de San Andres, La Paz, Bolivia.
2007 Master of Arts, Geography & Natural Resource Management, Graduated with Honors, San Francisco State University.
2013 Doctor of Philosophy, University of California, Merced

Publications:

Fernández, M., S. Blum, S. Reichle, Q. Guo, B. Holzman, and H. Hamilton. 2009. Locality uncertainty and the differential performance of four common niche-based modeling techniques. Biodiversity Informatics 6:36-52.

Fernández, M., D. Cole, W. R. Heyer, S. Reichle, and R. O. De Sa. 2009. Predicting Leptodactylus (Amphibia, Anura, Leptodactylidae) distributions: broad-ranging versus patchily distributed species using a presence-only environmental niche modeling technique. South American Journal of Herpetology 4:103-116.

Daszak, P., C. Zambrana-Torrelio, T. Bogich, **M. Fernández**, J. Epstein, K. Murray, and H. Hamilton. 2012. Interdisciplinary approaches to understanding disease emergence: The past, present, and future drivers of Nipah virus emergence. Proceedings of the National Academy of Sciences of the United States of America 110:3681-3688.

Fernández, M., H. Hamilton, O. Alvarez, and Q. Guo. 2012. Does adding multi-scale climatic variability improve our capacity to explain niche transferability in invasive species? Ecological Modelling 246:60-67.

Fernández, M., H. Hamilton, and L. M. Kueppers. 2013. Characterizing uncertainty in species distribution models derived from interpolated weather station data. Ecosphere in press.

ACKNOWLEDGEMENTS

This dissertation is the result of not only the five years it took to complete but the result of a research trajectory that began in 2003 when I was georeferencing South American amphibian specimens at the Smithsonian National Museum of Natural History. During this amazing journey through undiscovered lands I met fantastic people who provided me support, advice and guidance. Many of them walked by my side for short periods of time lighting the trail with a flashlight during particularly treacherous stages. Others joined the adventure later and skillfully pushed me through the finish line. A few stayed with me all the way. I am deeply grateful to each and every one of these people for sharing with me the opportunity to learn from them. Two persons that I do wish to single out for special recognition are my advisor Dr. Lara M. Kueppers for masterfully guiding my intellectual exploration, and Dr. Healy Hamilton for her wholehearted and never ending support. I am also forever thankful to Mariana and all my family, in space and time, for giving me the strength to pursue my enduring quest for learning.

ABSTRACT

CONSEQUENCES OF SPATIAL AND TEMPORAL CLIMATE VARIABILITY FOR SPECIES DISTRIBUTION MODELING

by

Miguel Alejandro Fernandez Trigoso

Doctor of Philosophy

University of California, Merced

Dr. Lara M. Kueppers, Chair

Our understanding of how species will respond to global change is still limited. Species distribution models (SDMs) are used to generate hypotheses regarding the potential distributions of species under different environmental conditions. However, that species observations and climatic variables are not measured at the same spatial and temporal resolution still hinders our ability to forecast species range shifts and expansions in response to global change. One of the possible consequences of this data mismatch is the observed discrepancy between realized climate niches in species' native and invasive ranges. In the first chapter of my dissertation, I address this issue by evaluating niche similarity between native and invaded ranges for 10 species using a combination of monthly and inter-annual climatic variability data. My results suggest that some species' ranges may be constrained by one aspect of climatic variability in the native range but a different one in the invaded range. A second issue, also a consequence of the spatio-temporal mismatch, is that weather station data are often spatially interpolated to match the species observations without any uncertainty assessment. The second chapter evaluates and quantifies the effects of three complementary aspects of uncertainty present in weather station data interpolations. I examine the influence of topographic heterogeneity, interannual climatic variability, and distance to weather station on SDM performance for 20 well observed North American breeding birds, and show that topographic heterogeneity has the highest contribution to omission errors, or false absences. A third consequence of the spatio-temporal data mismatch is the inability of global simulations to capture local manifestations of climate change. This inability can limit the capacity of SDMs to produce accurate simulations for species whose distributions depend on small-scale climate phenomena. While changes in global climatic patterns are projected using global climate model (GCM) simulations, local climatic trends are not always well represented by GCMs, or by simple downscaled projections derived from GCMs. In the final chapter of my dissertation, I use interpolated, fine-scale historic climate records in a novel approach to estimate the sensitivity of SDM's to locally coherent changes in temperature and precipitation at larger scales, using coast redwood (*Sequoia sempervirens*) as an example. Overall, the results of this thesis confirm the importance of establishing an appropriate relational basis in time and space between species and climatic observations. Historical collection records should be thoroughly analyzed and integrated with historical climatic time series to gain a better understanding of species' response to climate variability in the past, thereby informing the selection of appropriate spatio-temporal scales of climate variability for projections under present and future conditions.

INTRODUCTION

Our understanding of the fundamental unit of biogeography, species' geographic ranges (Angert, 2009), is derived from information on what organisms occur where in nature. Because available documented records of species occurrences are typically sparse in comparison to the complete distribution of a species, a variety of methods have been devised to visualize, infer and analyze species ranges based only on field samples (Fernández et al., 2009). These methods vary from simple convex hull plots around occurrence records to more sophisticated approaches that not only take advantage of the geographic position where species have been observed but also of the environmental variables that characterize these sites that can be used to deduce the conditions that are apparently conducive to species' survival.

Since the establishment of the Global Biodiversity Information Facility in 2001, there has been a substantial increase in the availability of species occurrence data as a result of large scale efforts to digitize and georeference specimens held in natural history museums (e.g., HerpNet, ORNIS), as well as efforts to improve online access to large observational data sources (e.g., BirdLife; Yesson et al., 2007). At the same time, information about climate, topography, soils, and photosynthetic activity have become available for almost the entire planet at increasingly fine spatial and temporal resolutions (Turner et al., 2003). Hardware and software coupled with the capacities to transfer, manipulate and analyze large quantities of digital information have allowed the development of geospatial tools that can relate species occurrences and environmental data in a more sophisticated way, providing estimates of potential suitable habitats. These estimates, once mapped onto the landscape, can result in practical and novel hypothesis of what structures species distributions in space and time (Franklin, 2009).

Species distribution models (SDMs), in their simplest form, are based on observations defining upper and lower bounds for temperature and precipitation within the species distribution range, resulting in what is known as the bioclimatic envelope for the species (Nix, 1986). More advanced statistical algorithms integrate multiple environmental variables and also account for the existing correlations among these variables (Phillips et al., 2006). These methods, embedded in a geographic information system (GIS) framework, combine taxonomic and geographic data describing organismal observations records with fine scale environmental variables to produce a set of correlative rules that identify the multidimensional environmental space in which the species was recorded (Peterson and Vieglais, 2001). This n-dimensional environmental space can then be projected onto geographic space, to estimate the likelihood that a location has favorable or unfavorable conditions for the species. The multidimensional environmental space also can be projected onto a scenario of future climate change to project where climatic conditions may be suitable for the species in the future. These approaches yield estimated distributions for large numbers of species for which we have relatively little physiological or demographic data from which to infer climate sensitivity.

There are many areas of environmental science where the demand for predictive models has outstripped the scientific understanding on which predictions are based (Baveye et al., 2009). Species distribution modeling has long been one of these areas (Huston, 2002). Modeling to increase our theoretical understanding of a system and modeling to carry out predictions for practical applications, are the two extremes in the continuum that represent the current application of SDMs. Inferential representations of species geographic ranges have become popular in many fields, from theoretical ecology and evolution (e.g., Lenoir et al., 2010; Thuiller et al., 2011) to applied conservation and resource management (de Oliveira et al., 2012; Hannah et al., 2013).

Despite their wide use, these practical summaries of biogeographic information are still controversial (Araujo and Peterson 2012; Saupe et al., 2012). Much of the criticism focuses on the simplifying assumptions (Wiens et al., 2009) necessary to reduce the full complexity of the system into a useful mathematical abstraction represented in a map while maintaining full scientific rigor. One key issue that has profound implications in the predictive power of the models and in the interpretation of the results is mixing data collected at different times and with different spatial resolution. This doctoral dissertation focuses on this issue and presents three case studies that illustrate how integrating species observations and climatic variables that are not measured at the same spatial and temporal scale hinders our ability to forecast species range shifts and expansions in response to global change. I also make recommendations, as explained in the following chapters, for deriving and interpreting SDMs with these issues in mind.

In the first chapter (Fernandez et al., 2012), I explore one of the possible consequences of not measuring species and climate observations at the same spatio-temporal scale. I hypothesize that multi-scale climatic variability is responsible for the lack of niche transferability often observed in invasive species. I create a novel spatial climate dataset, ClimVar, which is freely available at: http://ecoclim.org/. Using ClimVar, I assess environmental niche transferability using three different configurations of environmental layers: (1) monthly (2) inter-annual and (3) a combination of monthly and inter-annual climatic variability. I contrast SDMs between native and invaded ranges of a suite of taxonomically diverse and well documented invasive species. My results suggest that some species range limits might be constrained by one scale of climatic variability in the native range and a different one in the invaded range. If these findings are extrapolated to niche transferability in time, I suggest that historical collection records should be analyzed to understand species' response to multiple scales of climate variability in the past, thereby informing the selection of appropriate scales of climate variability in the future.

To match the spatial and temporal resolution of the species occurrences, weather station data often are interpolated and these resulting continuous data layers are incorporated into SDMs, often without any assessment of uncertainty. The second chapter of my dissertation (Fernandez et al., 2013) evaluates and quantifies the effects of three unrelated but complementary aspects of uncertainty present in weather station interpolations. I examine the influence of topographic heterogeneity, interannual variability, and distance

to weather station on SDM performance (over- and under-prediction errors). My findings show that these three metrics of uncertainty in interpolated weather station data have varying contributions to over- and under-prediction errors in SDMs. Topographic heterogeneity has the highest contribution to omission errors; the lowest contribution to commission errors is from Euclidean distance to weather station. The results confirm the importance of establishing an appropriate relational basis in time and space between species and climatic layers, providing key operational criteria for selection of species observations used in SDMs. My findings also highlight the importance of identifying weather station locations used in interpolated products, which will allow characterization of some aspects of uncertainty and identification of regions where users need to be particularly careful when making a decision based on a SDM.

SDM studies that evaluate the effect of climate change on terrestrial ecosystems often use scenarios from downscaled Global Climate Model outputs (GCMs). These simulations do not capture local manifestations of greenhouse gas-induced warming such as the hypothesized intensification of wind-driven coastal upwelling along the California coast (Bakun et al., 2010). Ignorance of this local effect reduces GCM's ability to provide accurate scenarios of future climate in coastal ecosystems. In the third chapter of my dissertation I take advantage of naturally occurring variability in the high-resolution historic climatic record, to develop multiple scenarios of coastal climate. I use these scenarios to estimate potential suitable habitat for coastal redwoods (*Sequoia sempervirens*) under "normal" combinations of temperature and precipitation, and under anomalous combinations, while maintaining coherent relationships between regional climate and coastal upwelling. By comparing historical state-wide conditions to GCM projections of state-wide climate from the CMIP5 archive, I found that warmer (but normal precipitation) historic scenarios are equivalent to changes projected for California for the 2020's and 2030's. Our results suggest that we could potentially use this high resolution "equivalent" as an alternative to downscaled GCMs in species distribution modeling and produce more reasonable and physically accurate estimates of the anticipated range shifts in response to projected climate change. However since none of the years in the past century provide mean scenarios beyond the 2030's, the approach is limited to near term projections.

CHAPTER 1
Does adding multi-scale climatic variability improve our capacity to explain niche transferability in invasive species?

Introduction

Our understanding of the fundamental unit of biogeography, species' geographic ranges (Angert, 2009), is derived from documented records of species occurrences in nature. Distributions maps have been inferred in many ways, from simply plotting convex hulls around occurrence points on a map (Burgman and Fox, 2003), to applying complex geostatistical tools that integrate specimen location data with fine scale environmental information, resulting in a species distribution model (SDM; for reviews see: Schröder, 2008; Elith and Leathwick, 2009; Zimmermann et al., 2010). Also known as bioclimatic models (e.g., Titeux et al., 2009) or ecological niche models (e.g., Hegel et al., 2010), SDMs have aided researchers in analyzing possible biogeographic scenarios and have provided valuable visualizations (Franklin and Miller, 2009). SDMs are now established as a key tool used in a variety of fields ranging from ecology and evolution (e.g., Graham et al., 2004; Lenoir et al., 2010), to natural resource management and conservation planning (e.g., Rodriguez et al., 2007; Thorn et al., 2009).

Despite their wide use and promise (Franklin and Miller, 2009; Drew et al., 2011), these practical summaries of biogeographic information are still controversial (Elith and Graham, 2009; Elith et al., 2010; Terribile et al., 2010; Barve et al., 2011; Rota et al., 2011). A critical aspect of SDMs is related to the niche concept developed by Hutchinson (1957); SDMs build a multivariate statistical representation of a species niche by relating species occurrences to environmental predictors. Applying SDMs to forecast species range shifts in response to climate change (e.g., Tingley et al., 2009; Carroll, 2010) or to predict invasive species range expansions (e.g., Roura-Pascual et al., 2009; Smolik et al., 2010), results in the same niche being projected into a different temporal or spatial context (i.e., niche transferability; (Phillips, 2008). These projections assume that the statistical and mechanistic relationships between species and environment remain static in the transferred context (e.g., Jeschke and Strayer, 2008; Steiner et al., 2008; Rödder and Lötters, 2009). However, studies focusing on invasive species have noted that while SDMs can perform well in areas for which the models have been fitted, they often fail in predicting occurrences of new invasions when projected into a different spatial context, showing a mismatch between realized climate niches in species' native and invasive ranges (Araujo and Rahbek, 2006; Randin et al., 2006; Broennimann et al., 2007; Beaumont et al., 2009; da Mata et al., 2010; Rödder and Lötters et al., 2010).

The choice of appropriate environmental data is another critical aspect of SDMs (Franklin and Miller, 2009) that requires additional attention (Suarez-Seoane et al., 2004; Peterson and Nakazawa, 2008). While past research efforts that address the issue of environmental predictors in SDMs have focused on increasing spatial resolution of

gridded datasets (e.g., Kriticos and Leriche, 2010), and testing alternative environmental variables (e.g., Synes and Osborne, 2011), there has been little attention to the various scales of variability inherent in the global climate system (Zimmermann et al., 2009). These natural fluctuations in climatic phenomena, such as temperature and precipitation, can range in a temporal scale from high (e.g., monthly) to low frequency (e.g., inter-annual), and are referred as multi-scale climatic variability (Diaz and Markgraf, 2000).

Recent studies suggest that the inclusion of these multi-scale measures of climatic variability can improve our understanding of species geographic limits (Jackson et al., 2009; Giesecke et al., 2010; Reside et al., 2010; Jiguet et al., 2011). However, these efforts, while important, do not provide clear guidelines for environmental variable selection for SDMs addressing questions related to modeling range expansions or shifts. In addition, opportunities to apply environmental predictors that characterize multi-scale climate variability in SDMs are limited by the lack of available global datasets offering temporal series at spatial resolutions relevant to biodiversity and its interaction with the environment (Heikkinen et al., 2006; Sax et al., 2007, Kremen et al., 2008). The best and most widely used approximation is Worldclim (Hijmans et al., 2005), an interpolated climatic dataset that characterizes monthly variability only consisting of mean, maximum and minimum temperatures, and total precipitation averaged for each month over a 50 year period, which fails to capture inter-annual climatic variability (Jackson et al., 2009).

In this study, we hypothesize that multi-scale climatic variability is responsible for the lack of transferability. The main purpose of this paper was not to generate a new gridded interpolated climatic dataset, but rather to examine the effect of multi-scale climatic variability on niche transferability (Phillips, 2008), using invasive species as a case study. Since a gridded characterization of inter-annual climatic variability was not previously available, we created a novel spatial climate dataset, ClimVar, which is freely available at: http://ecoclim.org/. Using ClimVar, we assess environmental niche transferability using three different configurations of environmental layers: (1) monthly (2) inter-annual and (3) a combination of monthly and inter-annual climatic variability. We contrast SDMs for native versus invaded ranges of a suite of taxonomically diverse and well documented invasive species, using Hellinger similarity index *I*. We ask if the inclusion of multi-scale climatic variability can shed light on whether or not species remain in the same n-dimensional climatic space when moved into a different spatial and temporal context. Our study indirectly contributes to understanding the magnitude of uncertainty in SDM applied to invasive species due to natural climatic variability.

Methods

Species and geographic occurrence data

We evaluated ecological niche transferability using a selected subset of terrestrial macrobiotic widespread organisms derived from the International Union for Conservation of Nature (IUCN) list of the top 100 Invasive Species (Lowe et al., 2004). Each of these species has at least 100 unique georeferenced localities in each of the native and invaded

ranges, precise to seconds in latitude and longitude. Most of the georeferenced locality information was obtained from the Global Biodiversity Information Facility. However, complementary datasets were also included (see Table 1).

Observation data points for each species were attributed to native or invaded ranges according to range discriminations described in the Global Invasive Species Database, thus creating two geospatial databases for each species, one for the native and one for the invaded range (*Appendix A*, Figs. S1-S10). Occurrence spatial bias (Hortal et al., 2008) was avoided by ensuring that the native range was sufficiently represented by available occurrence data, contrasting native range observations with known range descriptions. Point locality information and georeferenced range literature descriptions were transformed into raster format using the ArcGIS v10 and a grid of 50 km^2. The two resulting raster layers, for each species, were compared (i.e., one map was subtracted from the other) using raster calculator in ArcGIS v10 Spatial Analyst, and only species with ≥80% similarity between the two maps were selected for the analysis. Ten species of the initial list of 100 species fulfilled all criteria (i.e., terrestrial, macrobiotic, with at least 100 observations and with a good representation of the native range in the locality information), including six plants, two amphibians, one bird and one insect (Table 1).

Table 1. Species list with specimen georeferenced observation data source and region.

Species (common English name)	Group	Native	Invaded	Sources
Rhinella marina (Cane toad)	Amphibian	1919	729	GBIF, CAS, NMNH
Lantana camara (Spanish flag)	Plant	1118	640	GBIF, TNC
Leucaena leucocephala (White leadtree)	Plant	450	515	GBIF, TNC
Linepithema humile (Argentine ant)	Insect	265	1020	GBIF, CAS, ROURA
Mimosa pigra (Mimosa)	Plant	399	108	GBIF, TNC
Lithobates catesbeianus (American bullfrog)	Amphibian	4141	2187	GBIF, CAS, NMNH
Sphagneticola trilobata (Creeping ox-eye)	Plant	451	158	GBIF, TNC
Sturnus vulgaris (European starling)	Bird	176300	116700	GBIF
Tamarix ramosissima (Saltcedar)	Plant	3109	1173	GBIF
Ulex europaeus (Common gorse)	Plant	35241	479	GBIF

CAS California Academy of Sciences
GBIF Global Biodiversity Information Facility
NMNH Smithsonian Institution
ROURA Roura-Pascual et al. (2004)
TNC The Nature Conservancy

Environmental layers

Nineteen climate-derived environmental layers, at a resolution of 10 km^2, were obtained from Worldclim (Hijmans et al., 2005). These variables, commonly referred to as Bioclim layers (Table 2), are generated from interpolated global weather station data and represent biologically relevant aspects of temperature, precipitation and seasonality. This dataset has been widely utilized because of its global extent and high spatial resolution relative to similar products (e.g., Purvis et al., 2011; Roura-Pascual et al., 2011; Zelazowski et al., 2011). However, these climatic layers are available only as an averaged

product for the period of 1950 to 1999, providing only a single value for each climatic variable.

To provide a measure that quantifies inter-annual climatic variability, we generated an original gridded dataset using historical weather station observations obtained from two complementary sources: the World-wide Agro-climatic Database from the Food and Agriculture Organization (FAOCLIM-2, 2005), and the Global Historical Climatology Network (GHCN; Vose et al., 1992). Time series from both sources were combined in 28,032 stations for monthly average temperature and 47,416 stations for precipitation. Duplicates were removed following protocols from Hijmans et al. (2005), with stations selected that have been recording data for at least the last 30 years (e.g., 1970 to 2000), resulting in 23,012 stations for monthly average temperature and 39,631 stations for precipitation (*Appendix A*, Figs. S11 and S12). For each station, the coefficient of variation for monthly values of precipitation and standard deviation for mean monthly values of temperature were calculated across years. The latitudes and longitudes of the corresponding weather stations were interpolated to a spatial resolution of 10 km^2 with Ordinary Kriging using a spherical model and a search radius of 12 points in ArcGIS v10. The monthly values of the coefficient of variation and monthly values of the standard deviation were used as interpolated magnitudes. One of the bioclimatic layers was used as a mask to assure seamless integration with the Worldclim variables.

Evaluation of the interpolation was achieved by means of data partitioning and cross validation (Picard and Cook, 1984). The coefficient of variation for monthly values of precipitation and standard deviation for mean monthly values of temperature were divided randomly into ten sub-samples each. Following Guo et al. (2010), the interpolation algorithm was run ten times, each time withholding one of the ten sub-samples from the interpolation. The withheld data was then used to calculate the Root Mean Squared Errors (RMSE), a metric that evaluates accuracy of interpolated surfaces (Aguilar et al., 2005):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (x_m - x_0)^2}{n}}$$

where x_m is the interpolated model, x_o is the observation, and n is the total number of points (Figure 1).

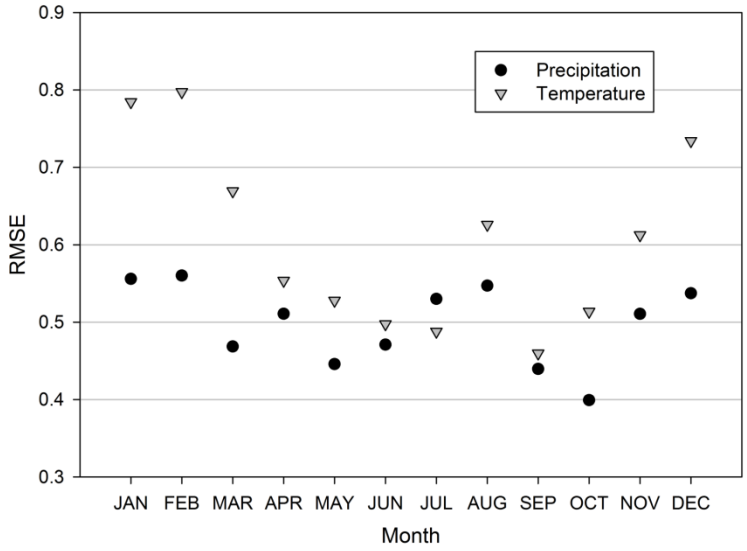

Figure 1. Interpolation accuracy evaluation. Root Mean Squared Errors (RMSE) results for precipitation and temperature based on a ten-fold cross validation.

The complete datasets of both the coefficient of variation and standard deviation points were used in the final interpolation, which resulted in 24 high-resolution novel variables that characterize monthly inter-annual climatic variability for precipitation and mean temperature at a global extent, referred to as ClimVar, and freely available for download at http://ecoclim.org/.

Based on Bioclim and ClimVar variables (Table 2), three configurations of climate layers were created: (1) Bioclim layers, representing monthly variability; (2) ClimVar layers, representing inter-annual variability; and (3) a combination of Bioclim and ClimVar layers, which represent monthly and inter-annual climatic variability.

To use ClimVar in a global test projecting species potential distributions from one hemisphere to the other, a reverse seasonality evaluation was required. We plotted and compared the following values against latitude: (1) mean monthly temperature, (2) total precipitation, (3) temperature inter-annual variability and (4) precipitation inter-annual variability for four months that represent the peak of each season in the northern hemisphere. The results indicate that the clear patterns of reverse seasonally observed in mean monthly temperature and total precipitation, don't exist in temperature inter-annual variability and precipitation inter-annual variability, which allow us to conclude that reverse seasonality is not an issue in ClimVar at global scale, nor at regional scale (*Appendix A*, Fig. S13), which support the use of the ClimVar dataset in combination with the 19 Bioclimatic variables without any transformations. To address the issues of multicolinearity of environmental variables (e.g. Peterson et al., 2007; Holt et al., 2009), each of the three variable configurations was subjected to a variable transformation, inside ArcGIS v10 (ESRI, 2010) using principal components analysis (PCA). PCA is a multivariate data reduction methodology previously employed in SDM (e.g., Cadena and

Loiselle, 2007; Drew et al., 2011; Fuller et al., 2011). Orthogonal components were retained by calculating the number of components with eigenvalues over one, which cumulatively explained most of the overall variation in the original variables (Everitt, 2005). For each climate layer configuration, 12 principal components were retained as the new predictor variables to be used in the SDMs.

Table 2. List of variables used in the analysis.

Variable name (units)	Dataset
Annual Mean Temperature (^{o}C)	Bioclim
Mean Diurnal Temp Range (^{o}C)	Bioclim
Isothermality	Bioclim
Temperature Seasonality	Bioclim
Max Temperature of Warmest Month (^{o}C)	Bioclim
Min Temperature of Coldest Month (^{o}C)	Bioclim
Temperature Annual Range (^{o}C)	Bioclim
Mean Temperature of Wettest Quarter (^{o}C)	Bioclim
Mean Temperature of Driest Quarter (oC)	Bioclim
Mean Temperature of Warmest Quarter (^{o}C)	Bioclim
Mean Temperature of Coldest Quarter (^{o}C)	Bioclim
Annual Precipitation (mm)	Bioclim
Precipitation of Wettest Month (mm)	Bioclim
Precipitation of Driest Month (mm)	Bioclim
Coefficient of variation for annual precipitation	Bioclim
Precipitation of Wettest Quarter (mm)	Bioclim
Precipitation of Driest Quarter (mm)	Bioclim
Precipitation of Warmest Quarter (mm)	Bioclim
Precipitation of Coldest Quarter (mm)	Bioclim
Standard deviation for January Mean Temperature (^{o}C)	Climvar
Coefficient of Variation for January Precipitation	Climvar
Standard deviation for February Mean Temperature (^{o}C)	Climvar
Coefficient of Variation for February Precipitation	Climvar
Standard deviation for March Mean Temperature (^{o}C)	Climvar
Coefficient of Variation for March Precipitation	Climvar
Standard deviation for April Mean Temperature (^{o}C)	Climvar
Coefficient of Variation for April Precipitation	Climvar
Standard deviation for May Mean Temperature (^{o}C)	Climvar
Coefficient of Variation for May Precipitation	Climvar
Standard deviation for June Mean Temperature (^{o}C)	Climvar
Coefficient of Variation for June Precipitation	Climvar
Standard deviation for July Mean Temperature (^{o}C)	Climvar
Coefficient of Variation for July Precipitation	Climvar
Standard deviation for August Mean Temperature (^{o}C)	Climvar
Coefficient of Variation for August Precipitation	Climvar
Standard deviation for September Mean Temperature (^{o}C)	Climvar
Coefficient of Variation for September Precipitation	Climvar
Standard deviation for October Mean Temperature (^{o}C)	Climvar
Coefficient of Variation for October Precipitation	Climvar
Standard deviation for November Mean Temperature (^{o}C)	Climvar
Coefficient of Variation for November Precipitation	Climvar

Species distribution modeling

We constructed the SDMs using the program MaxEnt v3.3.1 (Phillips et al., 2006; Phillips and Dudik, 2008; Elith et al., 2011). MaxEnt is a correlative niche model that uses the principle of maximum entropy to estimate a set of functions that relate environmental variables and species occurrence in order to approximate species' niche and potential geographic distribution (Phillips et al., 2006). We chose MaxEnt because of its proven performance with presence-only data (Elith et al., 2006; Elith and Leathwick, 2009), relative to alternative SDM techniques (but see Li et al., 2011).

For each species, native and invaded ranges were modeled separately (*Appendix A*, Figs S14-S23). Fifty percent of the occurrence points were withheld from the model to be used as independent test data. This process was repeated 100 times with randomly permuted occurrence samples to produce the underlying probability density. All other MaxEnt settings relating to model parameterization were left at their default values.

Model evaluation was performed using the area under the curve (AUC) of the receiver operating characteristic plot analysis (Fielding and Bell, 1997). AUC was chosen because it is a widely accepted, threshold-independent metric of SDM performance (Marmion et al., 2009; Warren et al., 2010) that provides an overall picture of how well the data fit the model (but see Lobo et al., 2008), and has previously been used in comprehensive SDM evaluations (Elith et al., 2006). Model predictive performance was assessed by comparing AUC scores among the three environmental layer configurations and across the ten species in their native and invaded ranges.

Niche similarity

To quantify environmental niche similarity between the species' native and invaded ranges, we used a measure, *I*, derived from Hellinger's similarity distance (Warren et al., 2008). Hellinger's *I* quantifies the difference between two normalized probability distributions (e.g. two SDM outputs) and ranges from 0, when species' niches do not overlap, to 1, where there is a complete overlap between species. This niche overlap metric was selected because it allows a threshold-independent comparison of two continuous raster layers. The *I* statistic is defined as:

$$I = 1 - \frac{1}{2}\sqrt{\sum(\sqrt{A} - \sqrt{B})^2}$$

where *A* and *B* represent the two normalized SDM outputs that are to be compared pixel by pixel. Statistical significance was assessed following Warren et al. (2008), repeating the comparison one hundred times, based on random samples with replacement of the original locality data from the native and invaded ranges. Each of these one hundred locality subsets for the native and invaded range was used as input in MaxEnt to produce a SDM. Once the models were created, ENMTools v1.0 (Warren et al., 2010) was used to calculate *I* for each new pair of maps and provide an average.

Assessing the statistical significance of the differences among niche similarity values and among SDM model performance values was accomplished with an ANOVA, using a multiple range test based on Fisher's least significant difference procedure in Statgraphics Centurion v16.1 (*Appendix A*, Table S1 and S2).

Results

The evaluation of SDM performance in both the native and invaded ranges based on the three environmental layer configurations produced three general patterns among species (Figure 2). For 60% of the species in the native range (*L. camara, L. humile, L. catesbeianus, R. marina, T. ramosissima* and *U. europaeus*), and 70% of the species in the invaded range (*L. leucochephala, L. humile, L. catesbeianus, R. marina, S. vulgaris, T. ramosissima* and *U. europeus*), SDMs showed significant improvement in the capacity to accurately predict the withheld test data when monthly and inter-annual climatic variability were used in combination. For 40% of the species in the native range (*L. leucocephala, M. pigra, S. trilobata* and *S. vulgaris*), and 30% of the species in the invaded range (*L. camara, M. pigra* and *S. trilobata*), SDMs showed no improvement in performance when multiple scales of variability were used over models that only considered monthly variability. For all cases, SDMs that were based on inter-annual climatic variability only, performed significantly worse than the other two layer configurations.

Contrasting native and invaded ranges for the SDM performance results (Figure 2) showed that for 70% of the species, the same layer configuration produced the highest-performing SDMs in the native and invaded range (*L. humile, L. catesbeianus, M. pigra, R. marina, S. trilobata, T. ramosissima* and *U. europaeus*). In contrast, *L. leucocephala* and *S. vulgaris* showed different results between the native and invaded ranges; no improvement in performance was found when multi-scale variability was used over models that only considered monthly variability in the native range, but a significant improvement was found when the combined layer configuration was used to produce the models in the invaded range. Conversely, for *L. camara*, native range SDMs did show improvement in model performance when the combined layer configuration was considered, however, the invaded range SDMs did not show difference in model performance between monthly and multi-scale variability.

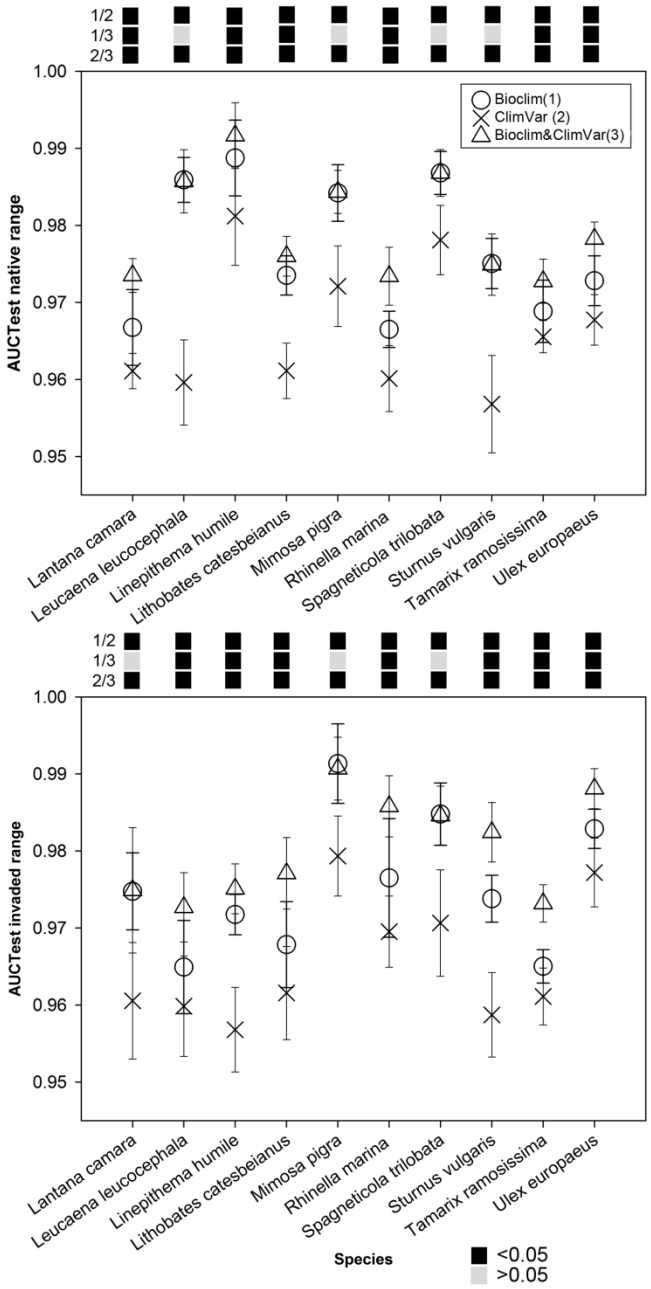

Figure 2. Comparison of the influence of three configurations of climatic datasets on MaxEnt model performance based on testing AUC values for 100 bootstrap models. Higher AUC values represent higher performing SDMs. Circles represent the performance of SDMs considering the monthly variability layer configuration (Bioclim); the cross represents the performance of SDMs considering inter-annual variability layer configuration (ClimVar); and the triangle represents the performance of SDMs considering both scales of climate variability (Bioclim+ClimVar). The predictive performance of MaxEnt based on ClimVar alone is relatively low for most species in the native and invaded range. There is an improvement in predictive performance when

Bioclim and ClimVar are used in combination (60% of the species in the native range and 70% of the species in the invaded range). For 30% of the species, the layer configuration that produced the best models in the native range was not the same that produced the best models in the invaded range. Black squares represent statistical significant differences (<0.05) between models in an ANOVA with a Fisher's least significant difference procedure where: 1/2: Bioclim over ClimVar; 1/3: Bioclim over Bioclim-ClimVar; 2/3: ClimVar over Bioclim-ClimVar.

The evaluation of native and invaded niche similarity (Figure 3) showed that for 90% of the species, the most similar niches were obtained when monthly variability only was considered. The exception was *S. vulgaris,* for which the highest similarity values between the native and invaded ranges were obtained when inter-annual variability only was considered. For *L. leucocephala* there was no significant difference in climatic niches between native and invaded ranges when monthly and inter-annual climate layer configurations were considered. Finally, our results showed that the combination of monthly and inter-annual climate variability did not produce the most similar results in any of the species.

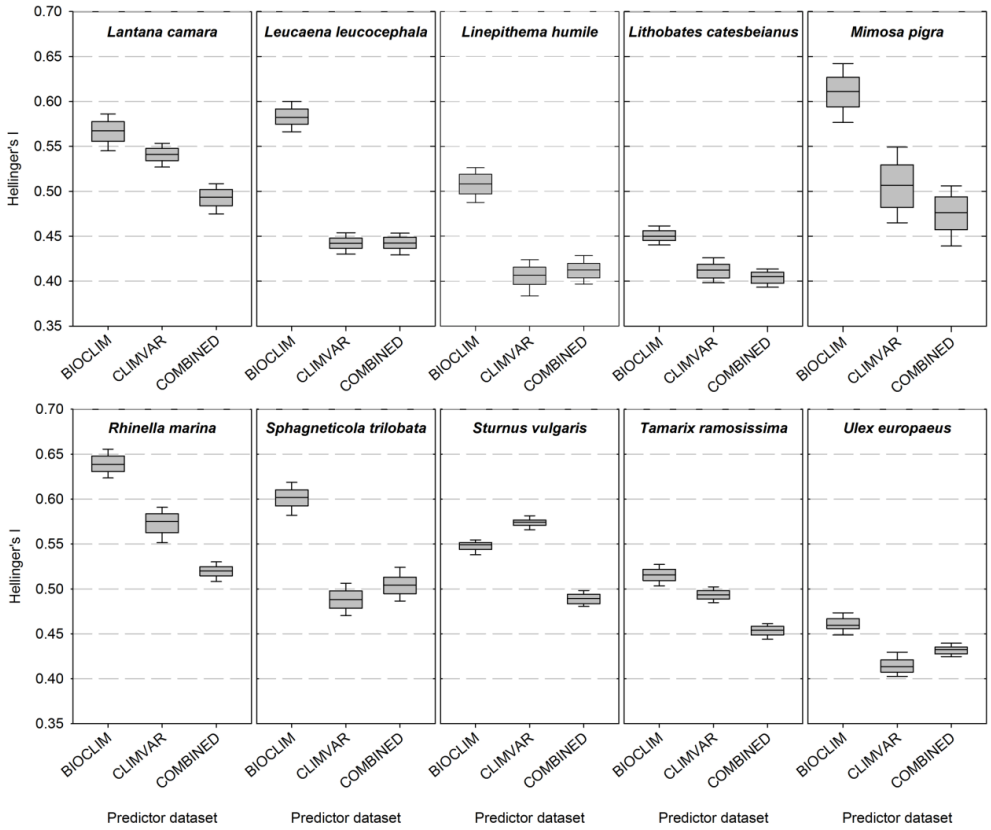

Figure 3. Comparison of values of niche similarity among the invaded range and the native range as defined by MaxEnt based on the three environmental layer configurations: monthly (Bioclim), inter-annual variability (ClimVar), and both scales of climate variability (Bioclim+ClimVar), based on 100 bootstrap models. A value of 1 for

the Hellinger's *I* index, represents a complete similarity between native and invaded range; a value of 0 represent a complete dissimilarity. For 90% of the species, native and invaded range comparison resulted in more similar SDMs when considering Bioclim only. For *S. vulgaris*, the most similar SDMs between native and invaded were obtained when ClimVar only was considered. ClimVar did not produce the most similar niches in any case.

Discussion

Global gridded datasets that characterize inter-annual climatic variability at spatial resolutions relevant to biodiversity do not exist. Currently available multi-year datasets that can be used to derive inter-annual variability either have low spatial resolution (CRU3.0; Mitchell and Jones, 2005) or the spatial extent is restricted to a region (PRISM; Daly et al., 2000). To test our hypotheses, we have derived a dataset that characterizes inter-annual climatic variability at 10km^2 for the terrestrial surface of the globe.

The ClimVar dataset is limited by the data and methods that were used to construct it. Weather station data is not evenly distributed in time or space, and a large fraction of meteorological data is still not digitized (Broennimann et al., 2008). Users should be aware that the inter-annual variability layers have relatively high degree of uncertainty in regions with low density of weather stations (Supplementary material, Figure S36 and S37). Yet, researchers interested in the spatio-temporal aspects of biodiversity can use these layers to move beyond monthly climatic means, and frame their hypotheses from a recent historical climatic perspective. Alternative global datasets that focus on climatic means, which are also based on interpolated weather station data (e.g., Worldclim, CliMond), use more sophisticated interpolation methods that account for co-variables in the regressions. Since the relationship between elevation and inter-annual climatic variability is not as clear as the relationship between elevation and temperature, we could not follow the same approach. For ecological niche modelers, ClimVar represents a way to improve model performance for species in which inter-annual climatic variability plays a factor in determining their distributions.

The idea that invasion success is based in part on the differential levels of ecological restrictions between native and invaded ranges is not new (Lockwood, 2009). However, global distributions of invasive species are still being projected under the assumption that species environmental preferences will be retained in the invaded area (Thuiller et al., 2005). By including an alternative scale of natural climatic variability in models of species that have already experienced major range expansions, our results have revealed a mismatch (i.e., the same set environmental layers did not produced the best models from native to invaded ranges) in realized climatic niches between the invaded and the native range for some of the species analyzed here. Although previous researchers have noted that SDMs trained in the native range often fail when projected into the invaded range (Randin et al., 2006; Broennimann et al., 2007; Beaumont et al., 2009; da Mata et al., 2010; Rödder and Lötters et al., 2010), our results demonstrate that some species' native ranges are being limited by one scale of climatic variability while the invaded ranges are

limited by a different one (*L. camara, L. leucocephala* and *S. vulgaris*). These results suggest that invasion success for these species can be attributed not only to the release from biotic interactions, but also to release from environmental limitations that existed in the native range. In other words, species in the native range had available suitable habitat in one scale of climatic variability but were being restricted in a different one, and when moved to the invaded range both scales allowed them to expand their range. These limitations might be a result of the multiple scales of climatic variability that affect natural systems, which suggests that their combined use may also improve our capacity to forecast species response to global change.

While our research results aim to improve understanding of invasion biology and advance the assessment of invasion risk, we could not find one general rule that applies to all the species and resolves the issue of niche transferability. One likely contributing factor is that many complexities of biological interactions in the species we analyzed remain unaccounted for. For example, the asymmetries we found between the native and invaded range for *L. camara* could be attributed to the hybridization process that this species has gone through in the invaded range with a closely related species, *Lantana depressa* (Caddotte et al 2006). Failure to differentiate the hybrid from *L. camara* may explain the disconnection between native and invaded range. Biological interactions also mediate invasion in *L. catesbeinaus,* where establishment has been facilitated in parts of the invaded range by the also invasive sunfish, which reduces the density of dragonfly larvae populations that attack the small *L. catesbeinaus* tadpoles (Cox, 2004). In our results, *L. catesbeinaus* shows no significant difference between native and invaded model performance. However, based on the mutualistic relationship that it has with the sunfish, it is clear that this biotic interaction is influential in determining at least a portion of its invaded range. Clearly, in these examples, climate is not the only factor that plays an important role in determining invasive establishment. However, when information on the ecology of the invasive species is limited, climate can provide a useful hypothesis of geographic potential in the absence of evolutionary changes and biotic interactions.

A limitation in our analysis results from treating all the species equally in the way they respond and perceive the environment. Clear differences exist in the dispersal abilities among the taxa that met our analysis criteria, and the spatial scales at which they operate. While wind-mediated dispersal is critical for some plant species, an amphibian's ability to disperse will mostly depend on the availability of water sources and the distance between them. Dispersal of invasive insects is often human-mediated, while bird dispersal ability is highly correlated with food availability. Future research on transferability in combination with the study of ecological traits at fine spatial scales is needed to increase our understanding on the subject.

In our approach, the native and invaded ranges were treated as homogeneous populations, decreasing the discriminative power needed to understand the effect of population structure in the species response to a highly variable spatio-temporal phenomenon such as climate (Lenoir et al., 2010). If a species that invades multiple regions originates from a single native population, we expect the more limited distribution of the originating

population will influence the range of conditions under which the species may establish in the invaded range. For example, *S. vulgaris* experienced a population bottleneck and exhibits very low levels of genetic variability, however, it still became a successful invasive (Lockwood, 2009). The decrease in genetic variability in the invaded range translates into a higher susceptibility to year-to-year environmental changes, (Cox, 2004), and might explain why the invaded range models in our results are more affected by the inclusion of inter-annual climatic variability. Future research that explores the idea of transferability in combination with molecular data, where finer scale traits are better correlated with particular geographies, might prove fruitful.

Overall, our results demonstrate that there were no improvements in the model predictions that use only inter-annual climatic variability. ClimVar, although unable to provide a full ecological picture by itself, does provide an improvement when combined with climatic means (e.g., Worldclim). Even with the addition of standardized measures of central tendency and dispersion (e.g., mean, standard deviation and coefficient of variation), we have yet to capture information about climatic extremes, which can have major impacts on biological systems. For example, the rapid spread of *L. camara* following the death of forest trees during severe drought in the southeast of North America (Mooney and Hobbs, 2000), and the major expansion of *M. pigra* in Australia following a major flood in Australia (Cox, 2004) show us how extreme and isolated events can cause stress in indigenous species and trigger episodic recruitment events. Thus a key question that remains to be answered is: What additional levels of natural climatic variability can and should be included to improve analyses aimed at the relationship of climate and species distributions? Time series records of species occurrences, such as natural history museum collections, natural resource monitoring programs, and long term ecological research efforts, are invaluable assets toward addressing this question. Only by understanding species' response to particular scales of climate variability over time can we be better prepared to make informed decisions in the selection of appropriate levels of climate variability in the future.

CHAPTER 2
Characterizing uncertainty in species distribution models derived from interpolated weather station data

Introduction

An attempt to understand species distributions, the fundamental unit of biogeography (Angert, 2009; Lomolino, 2010), has stimulated the development of tools to model the geographic distribution of organisms as a function of environmental factors. These models are used not only to understand distributions under contemporary environmental conditions, but also to predict whether or not a species might find suitable habitat outside the boundaries of its current distribution (e.g., Vaclavik and Meentemeyer, 2012) and to investigate the response of a species to projected future climates or reconstructed paleo-climates (e.g., Fordham et al., 2012; Stigall, 2012). Species distribution modeling, also known as ecological or environmental niche modeling, has undergone an exponential growth in popularity and applications in recent years (Elith and Leathwick, 2009) and is now a frequently used method in multiple fields such as ecology, evolution, conservation biology, epidemiology and agriculture.

As with any model, the output of species distribution models (SDMs) is dependent on the quality of data upon which they are built. Two key sources of SDM input data include: (1) species observations in nature, such as georeferenced point occurrences, and (2) environmental variables, such as high-resolution gridded climate layers. Multiple correlations between environmental parameters and known locations of species occurrences are constructed using geostatistical algorithms, defining an n-dimensional space that represents the climatic requirements of a species. This inferred multidimensional space can then be projected back into geographic space to produce a map of the species' potential distribution. While the quality of species observations and their effect on SDMs have been extensively documented (Soberon et al., 2000; Graham et al., 2008; Hortal et al., 2008; Loiselle et al., 2008; Fernández et al., 2009; Lobo et al., 2010; Feeley and Silman, 2011; Naimi et al., 2011), the deficiencies and biases of environmental variables have seldom been considered (Peterson and Nakazawa, 2008; McInerny and Purves, 2011; Synes and Osborne, 2011), despite the key role they play in the process of building, evaluating and calibrating SDMs.

SDM applications that project models to a space or time other than that from which they were created often use interpolations based on weather station data (e.g., Fernández et al., 2012; Gonzales et al., 2010), in which the spatial pattern of uncertainties is non-uniform and highly variable (Johnson et al., 2000), and as such, can lead to misinterpretation of spatial and temporal accuracy by users (Beale and Lennon, 2012). Worldclim (Hijmans et al., 2005), a gridded climatic dataset that consists of monthly mean temperature and precipitation values averaged over the 1950 to 1999 period at one km^2 spatial resolution, is one example of an interpolated dataset widely used in SDMs where the degree of

uncertainty associated with individual cell values for a particular climatic variable is influenced by three elements: (1) spatial variability, (2) temporal variability, and (3) the density of available observations (Zhang and Goodchild, 2002). The first two elements are considered to be intrinsic characteristics of the parameter to be estimated, while the last one is considered to be a characteristic attributable to the observation system. Although these three unrelated but complementary elements represent the core of uncertainty characterization of interpolations for weather station data, their effect on SDM performance has not been quantified.

Recent studies suggest that poor model performance, in part, can be attributed to high levels of uncertainty in the environmental data (Kriticos and Leriche, 2010; Beale and Lennon, 2012; Kamino et al., 2012); however the relationship between model omission and commission errors and the degree of uncertainty in the interpolated environmental input layers has not yet been addressed in the literature. The aim of this study is to determine whether SDM performance can be directly attributed to any of these three aspects of uncertainty. Explicitly accounting for the underlying uncertainty in the weather station interpolated data, we investigate three hypotheses: 1) SDM omission and commission errors are more often found in regions with high levels of spatial variability, 2) SDM performance is degraded by the mismatch between the scale of climatic variability used to create the model and the scale at which species distributions respond, and 3) SDM omission and commission errors are expected to be higher in regions with relatively low density of weather stations. These expectations are tested using a combination of high quality species occurrence data and novel gridded datasets that include estimates of environmental uncertainty. Our results suggest that biogeographers can benefit from increased attention to the variability and uncertainty in gridded spatial climate data when developing and applying SDMs.

Methods

Species occurrence data

We selected a subset of twenty species of birds from distribution data compiled by the North American Bird Breeding Survey (BBS; Sauer et al., 2006). The BBS dataset is a well-vetted, standardized, spatially balanced, long-term source of bird species occurrences (Sauer et al., 2006). Importantly, it is a close approximation to a true presence-absence, multi-species, observational dataset available at a continental scale. From the full list, twenty species were selected (Table 3) based on the criteria that their breeding distributions are largely determined by climate (John Dumbacher, personal communication, June 2011); increasing the likelihood their distributions can be reasonably modeled using primarily climate variables.

Table 3. Species list.

Species	Distribution comments
*Callipepla squamata**	Range fluctuates in response to variability in winter rainfall (Giuliano and Lutz, 1993).
*Dendragapus obscurus**	Prefers high lands in the winter. Occurs in pine and fir forest habitats from sea level to 3,600 m (Johnsgard, 1988).
*Tympanuchus cupido**	Extirpated from much of the range in US. Native prairie is preferred, but also adapted to cropland (Schroeder and Braun, 1993).
*Centrocercus urophasianus**	Adapted to winter extremes. Distribution is reduced as a result of loss of sagebrush habitat. Current distribution estimated at 56% of pre-settlement (Drut et al., 1994).
Columba fasciata	Moves seasonally to areas higher or lower than normal range. Timing of breeding a factor of food availability (Howell and Webb, 1995).
*Buteo regalis**	Distribution and density closely associated with cycles of prey abundance determined by climate (NatureServe, 2012).
*Picoides borealis **	Cooperative breeder influenced by loss of habitat, requires >80 ha of continuous habitat. Dependent of fire-maintained, old-growth pine forest (BirdLife, 2012).
*Picoides nuttallii **	Endemic species to California and Baja California. Confined to oak woodlands (NatureServe, 2012).
*Picoides albolarvatus**	Fire suppression and fragmentation has contributed to range decline in the northern part of the distribution. Biology and ecology remains unstudied (IUCN, 2012).
*Melanerpes lewis **	Strongly associated with fire-maintained old-growth ponderosa pine (Saab and Dudley, 1998).
*Calypte anna ***	Moves to low elevations in the winter. The only hummingbird that spends the winter in northern climates (Johnsgard, 1983).
*Selasphorus platycercus ***	Some individuals have moved into urban and suburban areas of southwestern due to hummingbird feeders (Calder, 1994).
Selasphorus sasin	Apparent expansion in breeding range due to availability of non-native flowers (Johnsgard, 1983).
*Pyrocephalus rubinus **	Northern populations move south in the winter. Can be found between 0 and 3,000 m in elevation (NatureServe, 2012).
*Aphelocoma californica**	Can be found in scrub-brush, boreal forests and temperate forests. Well adapted to suburban areas (NatureServe, 2012).
*Calamospiza melanocorys ***	Arrives until late May to the northern edge of its range (NatureServe, 2012).
Limnothlypis swainsonii	Summer and winter distribution. One of the least observed of North American birds if it weren't for its loud song (BirdLife, 2012).
*Vermivora luciae ***	Incomplete information on breeding ecology. Arrives and departs early from breeding grounds perhaps to evade much of the summer heat (BirdLife, 2012).
*Vermivora virginiae ***	Limited information on distribution and habitat preferences (NatureServe, 2012).
*Dendroica caerulescens ***	Male common in forest at lower to mid-elevations, female uses shrubbier habitat at higher elevations. Mortality from exposure to cold or rainy weather (BirdLife, 2012).

* Year round distribution ** Winter and breeding distribution

For each species, observation data points representing multi-year survey routes (Sauer et al., 2003) were split into presence and absence. In order to avoid a subjective decision in the placement of the break between presence vs. absence across all survey years, two

complementary approaches were applied. First, species observations in transect location maps were plotted as histograms, supporting the detection of naturally occurring breaks in the data. Secondly, BBS range maps were compared to an independent source of species range descriptions (BirdLife International and Natureserve, 2012; *Appendix B*, Figs. S24-S43). Following Fernández et al., (2012), the independent range maps were transformed into a raster format that matched the BBS map's spatial resolution. The maps were compared using ArcGIS Version 10 Spatial Analyst, looking for the value in the classification of each BBS continuous map that provided the highest value of similarity among them. The two approaches agreed for all species. We therefore defined presences as a survey route location point along which a particular bird species had been recorded during at least one of the ten years that the route was visited (1994 to 2003); routes without at least one positive record were considered absences.

Climatic gridded data

Nineteen climate layers, at a resolution of 1 km^2, were obtained from Worldclim Version 1.4 (Hijmans et al., 2005). These variables, commonly referred to as bioclimatic layers (Table 2), represent biologically relevant aspects of temperature and precipitation. This gridded climatic dataset, which provides one of the finest spatial resolutions relative to other similar products at a global extent (e.g., Purvis et al. 2011, Roura-Pascual et al. 2011, Zelazowski et al. 2011), was chosen for multiple reasons. First, for North America, Worldclim was generated from interpolated weather station data obtained from the Global Historical Climatology Network (GHCN, Vose et al. 1992), and the World-wide Agro-climatic Database from the Food and Agriculture Organization (FAOCLIM-2 2005), datasets to which we also had access. Second, this climatic dataset is available only as an averaged product for the period of 1950 to 1999 and, therefore, does not account for interannual climatic variability. Third, Worldclim does not provide the user with an assessment of the quality of the information or uncertainty characterization in the data. Finally, this climatic dataset has been cited 1,534 different times since its publication (ISI, 2012), and constitutes a critical resource for studies in diverse scientific fields including ecology, conservation, paleobiology, public health, anthropology and developmental biology (e.g., de Oliveira et al. 2012, Levsen et al. 2012, Daszak et al. 2012, Kamilar et al. 2012, Rosell et al. 2012).

Uncertainty layers

To test our three hypotheses, we generated three gridded datasets that represent three different aspects of uncertainty in interpolated climatic data layers. The first dataset provided a metric of spatial variability, a factor known to contribute to biodiversity at the landscape level (Kreft and Jetz 2007). Spatial variability was quantified as topographic heterogeneity, measured by the number of unique elevation values within 25 km^2 of the target pixel derived from the one km^2 spatial resolution Shuttle Radar Topographic Mission digital elevation model (SRTM-DEM, Farr et al. 2007), using a Python script to iterate ArcGIS Version 10 Zonal Statistics tool.

The second uncertainty dataset was an index of temporal climate variation, ClimVar (Fernández et al. 2012), calculated for weather stations that had a record of at least 30 years. This dataset was chosen because it represents a fine resolution spatial characterization of interannual climatic variability, which represents the largest temporal fluctuation in the climatic system, as compared to daily, intraseasonal and interdecadal variability (Ghil 2002). ClimVar is based on the same combined sources of weather stations (Table 4) used to create Worldclim. The specific ClimVar layers used here are one standard deviation of mean annual temperature and the coefficient of variation of annual total precipitation.

The third dataset provided a measure of the density of available information for interpolation. This dataset was based on the 6,499 stations recording monthly average temperature and 8,671 stations recording precipitation for the continental United States (*Appendix B*, Fig. S44), the same weather station data used to produce ClimVar and WorldClim. Using the SRTM-DEM as reference for cell size, cell center position, and elevation value, we calculated a new gridded layer where each cell value reflected the combination of vertical and horizontal distance (i.e., Euclidean distance) from the center of the cell to the closest temperature or precipitation weather station (these were rarely equivalent). The Marine Geospatial Ecology Tools Version 0.8a44 and the Proximity Toolset in ArcGIS Version10 were used to determine the closest weather station and measure the distances as follows:

$$d = \sqrt{\left(x_2 - x_1\right)^2 + \left(y_2 - y_1\right)^2 + \left(z_2 - z_1\right)^2}$$

Where x_2- x_1 represents the longitudinal distance between any point in the reference shapefile (x_1) and the nearest weather station (x_2), y_2- y_1 represents the latitudinal distance between any point in the reference shapefile (y_1) and the nearest weather station (y_2), and z_2- z_1 represents the difference in elevation from any point in the reference shapefile (z_1) and the elevation at the nearest weather station (z_2).

Table 4. List of variables used in the analysis.

Code	Variable description (units)	Source
Bioclimatic layer	Annual Mean Temperature (^{o}C)	1
Bioclimatic layer	Mean Diurnal Temp Range (^{o}C)	1
Bioclimatic layer	Isothermality	1
Bioclimatic layer	Temperature Seasonality	1
Bioclimatic layer	Max Temperature of Warmest Month (^{o}C)	1
Bioclimatic layer	Min Temperature of Coldest Month (^{o}C)	1
Bioclimatic layer	Temperature Annual Range (^{o}C)	1
Bioclimatic layer	Mean Temperature of Wettest Quarter (^{o}C)	1
Bioclimatic layer	Mean Temperature of Driest Quarter (^{o}C)	1
Bioclimatic layer	Mean Temperature of Warmest Quarter (^{o}C)	1
Bioclimatic layer	Mean Temperature of Coldest Quarter (^{o}C)	1
Bioclimatic layer	Annual Precipitation (mm)	1
Bioclimatic layer	Precipitation of Wettest Month (mm)	1
Bioclimatic layer	Precipitation of Driest Month (mm)	1
Bioclimatic layer	Coefficient of variation for annual precipitation	1
Bioclimatic layer	Precipitation of Wettest Quarter (mm)	1

Bioclimatic layer	Precipitation of Driest Quarter (mm)	1
Bioclimatic layer	Precipitation of Warmest Quarter (mm)	1
Bioclimatic layer	Precipitation of Coldest Quarter (mm)	1
Uncertainty layer	Topographic heterogeneity	2
Uncertainty layer	Standard deviation from mean annual temperature	3
Uncertainty layer	Coeficient of variation from annual total precipitation	3
Uncertainty layer	Euclidean distance to closest temperature weather station	2, 4 & 5
Uncertainty layer	Euclidean distance to closest precipitation weather station	2, 4 & 5

[1]Worldclim [2]SRTM [3]ClimVar
[4]FAOCLIM-2 [5]GHCN

Species distribution modeling

Continuous SDMs were generated using MaxEnt Version 3.3.3e (Phillips et al. 2006), a machine learning algorithm that uses the principle of maximum entropy to derive a set of rules correlating environmental variables and species occurrences to estimate the potential geographic distribution of a target species. MaxEnt was chosen because of its well established performance relative to alternative niche modeling techniques (Elith et al. 2006, Elith and Leathwick 2009, but see Li et al. 2011), and its capacity to deal with multicollinearity in the environmental variables, by considering redundant information without penalizing models by overfitting (Phillips et al. 2006, Phillips and Dudik 2008, Elith et al. 2011).

In order to avoid spatial autocorrelation (Peterson et al. 2011), all the localities that included presences and absences for each species were split into three subsets: 60% for model training, 20% for threshold selection and 20% for model evaluation (Figure 4) based on spatially structured partitioning and random selection of the data (modified from Daszsak et al. 2012). Only the presence portion of the 60% subset was used to train each SDM. The first 20% subset, which included presences and absences, was used to select the threshold value applied to convert the continuous SDM output into a binary map. A cutoff value that maximizes sensitivity and specificity was computed based on the area under the receiver operating characteristic (ROC) curves (Fielding and Bell 1997) using SigmaPlot Version 11.0. The remaining 20% subset, which also included presences and absences, was contrasted with the resulting MaxEnt binary map using a Python script to detect observations that were over- or under-predicted by the model. This process was repeated one hundred times, each time randomly sampling with replacement a new combination of 60-20-20. This analysis produced two combined datasets that include all of the localities that were identified as false positives (commission errors) or false negatives (omission errors) by at least one of the bootstrapped SDMs for each species. These combined datasets were then used as the reference locations for which associated values of uncertainty were extracted (Figure 4), using the Geospatial Modeling Environment Version 0.7.2 RC2 Isectpntrst tool (Beyer 2012).

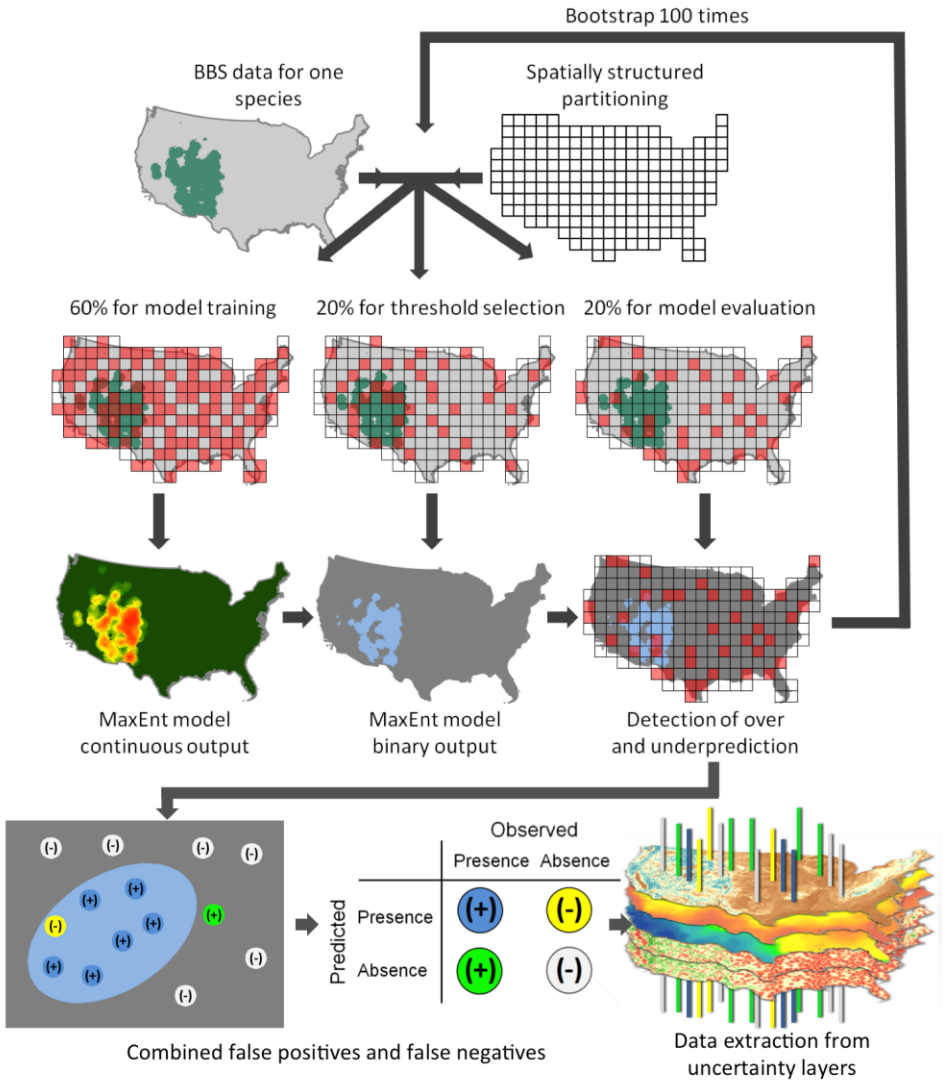

Figure 4. Flowchart describing spatially structured partitioning of the data and sampling strategy. All localities were randomly split into three subsets: 60% for model training, 20% for threshold selection and 20% for model evaluation. The presence portion of the 60% subset was used to train a SDM. The first 20% subset was used to select the threshold. Cutoff values that maximize sensitivity and specificity were computed based on the area under the ROC curve. The second 20% subset was contrasted with the resulting MaxEnt output to detect observations that were over- or under-predicted by the model. This process was repeated for each species to produce 100 bootstrapped iterations. The results were two combined datasets that include all the localities that were identified as false positives or false negatives by at least one of the bootstrapped SDMs for each species. These datasets were then used as the reference locations for which associated values of uncertainty were extracted.

Data analysis

To understand the relationship between performance errors in the SDMs and the three sets of layers that represent different aspects of uncertainty in the interpolated station data, we compared the uncertainty values associated with over- and under-predictions to a null distribution using the two sample Kolmogorov-Smirnov test (K-S) with Stata Version 11.2. The two sample K-S test is commonly used to assess whether two independent samples come from an identical distribution, making no assumptions about the normality of the data. The two sample K-S statistic *D* represents a measure of the maximum difference between the cumulative distribution functions for each sample (Conover 1999). For each species' withheld evaluation data, two different sets of 100 localities were randomly resampled with replacement (bootstrap, Efron 1982) 10,000 times. Using the two sample K-S test, their associated values of uncertainty were compared to produce the underlying probability density for *D* under the null distribution. From this distribution the mean value, D_{null}, was calculated (Figure 5a).

To compare the correctly predicted absences to the incorrectly predicted absences (i.e., false negatives), two subsets of 100 localities obtained from the withheld evaluation data were contrasted using the two sample K-S test. The process was bootstrapped 10,000 times to obtain the underlying probability density for *D*, in the case of the commission errors. From this distribution the mean value, D_{ov}, was calculated (Figure 5b).

To compare the correctly predicted presences to the incorrectly predicted presences (i.e., false positives), two subsets of 100 localities obtained from the withheld evaluation data were contrasted using the two sample K-S test. The process was also bootstrapped 10,000 times to obtain the underlying probability density for *D*, this time for the omission errors. From this distribution the mean value, D_{un}, was calculated (Figure 5c).

For each distribution, the mean value calculated for the resulting bootstrapped parameter (i.e., D_{ov} and D_{un}) was subtracted from the mean value calculated from its corresponding null distribution (i.e., D_{null}). The absolute magnitudes of the differences were used to determine the strength of the association between commission or omission errors and each of the uncertainty layers (Figure 5, Table 5, 6 and 7). The same approach was repeated for each uncertainty layer and each species, totaling 200 comparisons. Finally, confidence intervals were defined for *D* under the null distribution, the commission errors, and the omission errors, based on the percentile method (Dixon 1993), where the intervals were calculated and compared directly from the frequency distributions of the bootstrapped statistics ($Q2.5\%$ and $Q97.5\%$) to define statistical significance.

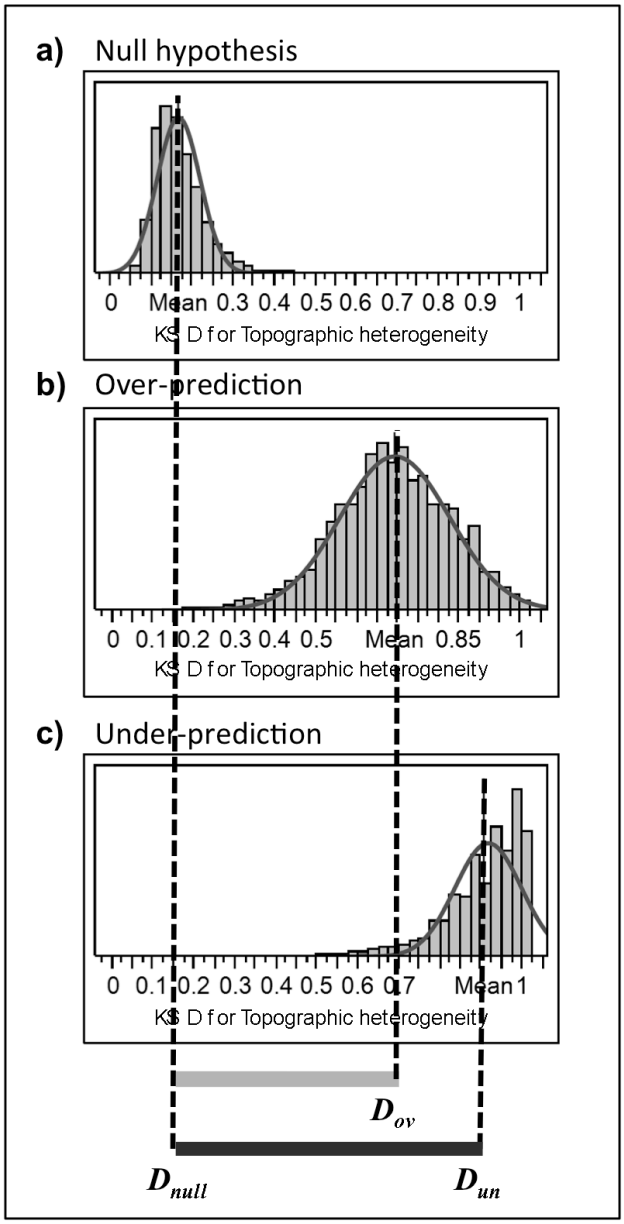

Figure 5. Normalized mean differences were obtained by subtracting the average value of the 10,000 bootstrap null distribution (D_{null}) for the Kolmogorov-Smirnov test (D) from the mean value from the distributions obtained by bootstrapping the over-predicted (X_{ov}) localities and the under-predicted (X_{un}) localities for *Vermivora luciae* (shown) and for all other species.

Table 5. Two sample Kolmogorov-Smirnov test results for topographic heterogeneity.

	Topographic heterogeneity									
	Overprediction					Underprediction				
Species name	Mean Obs. Coef.	Boot. Std. Err.	Q (2.5%)	Q (97.5%)	Sig. 0.05	Mean Obs. Coef.	Boot. Std. Err.	Q (2.5%)	Q (97.5%)	Sig. 0.05
A. californica	0.49	0.13	0.24	0.79	-	0.87	0.12	0.53	1	*
B. regalis	0.55	0.16	0.25	0.89	-	0.80	0.18	0.40	1	*
C. melanocorys	0.57	0.19	0.25	0.95	-	0.45	0.14	0.22	0.81	-
C. squamata	0.49	0.17	0.22	0.90	-	0.79	0.19	0.37	1	*
C. anna	0.73	0.10	0.50	0.93	*	0.90	0.12	0.50	1	*
C. urophasianus	0.56	0.12	0.35	0.81	*	0.87	0.14	0.45	1	*
C. fasciata	0.68	0.10	0.46	0.88	*	0.86	0.15	0.45	1	*
D. obscurus	0.64	0.09	0.45	0.83	*	0.85	0.16	0.44	1	*
D. caerulescens	0.49	0.13	0.26	0.81	-	0.77	0.20	0.37	1	*
L. swainsonii	0.51	0.18	0.24	0.94	-	0.85	0.15	0.49	1	*
M. lewis	0.55	0.11	0.34	0.78	*	0.93	0.08	0.65	1	*
P. albolarvatus	0.72	0.11	0.49	0.91	*	0.89	0.13	0.48	1	*
P. borealis	0.57	0.17	0.27	0.93	-	0.87	0.15	0.45	1	*
P. nuttallii	0.75	0.11	0.50	0.94	*	0.85	0.17	0.43	1	*
P. rubinus	0.31	0.11	0.15	0.59	-	0.91	0.10	0.60	1	*
S. platycercus	0.55	0.12	0.33	0.79	*	0.88	0.12	0.51	1	*
S. sasin	0.77	0.13	0.47	0.99	*	0.93	0.06	0.77	1	*
T. cupido	0.49	0.13	0.25	0.78	-	0.79	0.20	0.38	1	*
V. luciae	0.69	0.13	0.41	0.94	*	0.91	0.08	0.70	1	*
V. virginiae	0.45	0.11	0.22	0.69	-	0.89	0.12	0.55	1	*

Table 6. Two sample Kolmogorov-Smirnov test results for interannual variability.

		Interannual variability									
		Overprediction					Underprediction				
Species name	Var.	Mean Obs. Coef.	Boot. Std. Err.	Q (2.5%)	Q (97.5%)	Sig. 0.05	Mean Obs. Coef.	Boot. Std. Err.	Q (2.5%)	Q (97.5%)	Sig. 0.05
A. californica	PPT	0.66	0.12	0.37	0.91	*	0.67	0.19	0.31	1.00	*
	T	0.56	0.11	0.33	0.77	*	0.68	0.19	0.32	0.99	*
B. regalis	PPT	0.52	0.13	0.28	0.82	-	0.61	0.19	0.29	0.99	*
	T	0.39	0.14	0.17	0.72	-	0.60	0.19	0.26	0.97	-
C. melanocorys	PPT	0.57	0.12	0.33	0.88	*	0.80	0.08	0.61	0.92	*
	T	0.45	0.15	0.23	0.81	-	0.61	0.12	0.39	0.88	*
C. squamata	PPT	0.69	0.14	0.36	0.93	*	0.59	0.19	0.27	0.98	-
	T	0.54	0.15	0.26	0.85	-	0.61	0.20	0.28	0.98	-
C. anna	PPT	0.79	0.16	0.41	0.99	*	0.61	0.17	0.29	0.99	*
	T	0.68	0.13	0.39	0.95	*	0.65	0.19	0.29	0.99	*
C. urophasianus	PPT	0.55	0.11	0.33	0.78	*	0.60	0.19	0.28	0.99	-
	T	0.37	0.12	0.17	0.64	-	0.60	0.19	0.02	0.98	-
C. fasciata	PPT	0.58	0.14	0.29	0.88	*	0.58	0.18	0.27	0.98	-
	T	0.59	0.09	0.39	0.75	*	0.60	0.20	0.27	0.99	-
D. obscurus	PPT	0.44	0.12	0.22	0.72	-	0.60	0.19	0.28	0.99	-
	T	0.34	0.10	0.16	0.57	-	0.60	0.19	0.27	0.99	-
D. caerulescens	PPT	0.65	0.13	0.38	0.88	*	0.69	0.18	0.32	1.00	*

	T	0.51	0.10	0.33	0.73	*	0.65	0.18	0.31	0.99	*
L. swainsonii	PPT	0.49	0.16	0.24	0.81	-	0.64	0.17	0.31	0.99	*
	T	0.73	0.10	0.45	0.92	*	0.70	0.19	0.31	0.99	*
M. lewis	PPT	0.56	0.13	0.30	0.82	*	0.60	0.19	0.28	0.99	-
	T	0.39	0.11	0.19	0.62	-	0.62	0.20	0.28	0.99	-
P. albolarvatus	PPT	0.60	0.14	0.33	0.91	*	0.62	0.18	0.29	0.99	*
	T	0.56	0.11	0.33	0.76	*	0.65	0.19	0.30	0.99	*
P. borealis	PPT	0.43	0.10	0.26	0.68	-	0.61	0.18	0.29	0.99	*
	T	0.78	0.12	0.48	0.94	*	0.64	0.19	0.29	0.99	*
P. nuttallii	PPT	0.85	0.11	0.57	1.00	*	0.60	0.17	0.29	0.99	*
	T	0.66	0.11	0.38	0.88	*	0.63	0.19	0.29	0.98	*
P. rubinus	PPT	0.57	0.15	0.28	0.90	-	0.69	0.19	0.33	1.00	*
	T	0.77	0.06	0.64	0.90	*	0.72	0.19	0.34	0.99	*
S. platycercus	PPT	0.61	0.12	0.35	0.85	*	0.62	0.19	0.29	0.99	*
	T	0.44	0.12	0.21	0.69	-	0.63	0.20	0.29	0.99	*
S. sasin	PPT	0.83	0.16	0.45	1.00	*	0.65	0.19	0.31	1.00	*
	T	0.75	0.14	0.37	0.97	*	0.67	0.20	0.31	0.99	*
T. cupido	PPT	0.39	0.12	0.20	0.70	-	0.61	0.18	0.29	0.99	*
	T	0.41	0.09	0.26	0.62	-	0.62	0.18	0.29	0.99	*
V. luciae	PPT	0.87	0.12	0.54	1.00	*	0.71	0.20	0.33	1.00	*
	T	0.73	0.14	0.40	0.96	*	0.70	0.20	0.32	0.99	*
V. virginiae	PPT	0.70	0.08	0.53	0.86	*	0.66	0.18	0.33	0.99	*
	T	0.38	0.11	0.17	0.62	-	0.68	0.19	0.32	0.99	*

Table 7. Two sample Kolmogorov-Smirnov test results for Euclidean distance.

		Euclidean distance									
		Overprediction					Underprediction				
Species name	Var.	Mean Obs. Coef.	Boot. Std. Err.	Q (2.5%)	Q (97.5%)	Sig. 0.05	Mean Obs. Coef.	Boot. Std. Err.	Q (2.5%)	Q (97.5%)	Sig. 0.05
A. californica	PPT	0.339	0.109	0.176	0.594	-	0.937	0.191	0.430	1.000	*
	T	0.341	0.108	0.178	0.592	-	0.841	0.187	0.438	1.000	*
B. regalis	PPT	0.362	0.129	0.180	0.677	-	0.863	0.187	0.459	1.000	*
	T	0.361	0.130	0.184	0.680	-	0.833	0.199	0.418	1.000	*
C. melanocorys	PPT	0.456	0.171	0.212	0.889	-	0.455	0.156	0.216	0.798	-
	T	0.452	0.170	0.255	0.859	-	0.441	0.147	0.217	0.773	-
C. squamata	PPT	0.429	0.165	0.200	0.888	-	0.946	0.117	0.546	1.000	*
	T	0.426	0.165	0.200	0.857	-	0.952	0.106	0.592	1.000	*
C. anna	PPT	0.444	0.153	0.212	0.814	-	0.963	0.105	0.576	1.000	*
	T	0.439	0.152	0.211	0.808	-	0.951	0.119	0.525	1.000	*
C. urophasianus	PPT	0.298	0.099	0.152	0.532	-	0.947	0.107	0.566	1.000	*
	T	0.300	0.099	0.150	0.527	-	0.919	0.129	0.515	1.000	*
C. fasciata	PPT	0.347	0.115	0.168	0.602	-	0.912	0.147	0.512	1.000	*
	T	0.356	0.117	0.170	0.613	-	0.901	0.156	0.500	1.000	*
D. obscurus	PPT	0.298	0.096	0.147	0.515	-	0.827	0.199	0.423	1.000	*
	T	0.300	0.095	0.149	0.513	-	0.834	0.194	0.429	1.000	*
D. caerulescens	PPT	0.594	0.123	0.362	0.857	*	0.792	0.216	0.378	1.000	*
	T	0.602	0.125	0.365	0.866	*	0.808	0.212	0.418	1.000	*
L. swainsonii	PPT	0.641	0.164	0.333	0.960	*	0.838	0.201	0.490	1.000	*
	T	0.638	0.163	0.326	0.948	*	0.854	0.198	0.500	1.000	*
M. lewis	PPT	0.291	0.091	0.152	0.505	-	0.957	0.099	0.606	1.000	*
	T	0.293	0.091	0.153	0.502	-	0.930	0.128	0.531	1.000	*

P. albolarvatus	PPT	0.402	0.127	0.202	0.691	-	0.991	0.043	0.898	1.000	*
	T	0.389	0.126	0.189	0.674	-	0.951	0.117	0.531	1.000	*
P. borealis	PPT	0.617	0.147	0.333	0.899	*	0.969	0.097	0.586	1.000	*
	T	0.609	0.152	0.330	0.908	*	0.972	0.088	0.667	1.000	*
P. nuttallii	PPT	0.479	0.158	0.232	0.857	-	0.983	0.060	0.776	1.000	*
	T	0.476	0.156	0.237	0.838	-	0.969	0.086	0.667	1.000	*
P. rubinus	PPT	0.423	0.130	0.197	0.695	-	0.913	0.162	0.500	1.000	*
	T	0.408	0.126	0.190	0.677	-	0.921	0.155	0.500	1.000	*
S. platycercus	PPT	0.300	0.098	0.150	0.526	-	0.895	0.171	0.500	1.000	*
	T	0.308	0.100	0.158	0.537	-	0.863	0.187	0.485	1.000	*
S. sasin	PPT	0.562	0.189	0.260	1.000	-	0.960	0.102	0.592	1.000	*
	T	0.563	0.175	0.296	1.000	*	0.971	0.080	0.667	1.000	*
T. cupido	PPT	0.584	0.127	0.333	0.837	*	0.871	0.192	0.480	1.000	*
	T	0.580	0.127	0.333	0.837	*	0.889	0.184	0.490	1.000	*
V. luciae	PPT	0.462	0.157	0.227	0.869	-	0.894	0.159	0.500	1.000	*
	T	0.463	0.156	0.233	0.847	-	0.936	0.111	0.586	1.000	*
V. virginiae	PPT	0.274	0.088	0.142	0.479	-	0.858	0.193	0.480	1.000	*
	T	0.280	0.089	0.143	0.484	-	0.873	0.187	0.500	1.000	*

Results

The results of our analysis indicate that false negatives and false positives, in that order, were significantly ($p<0.05$) associated (a) with regions of high topographic heterogeneity for 95% and 50% of the species (Figure 6a), (b) with regions of high interannual precipitation variability for 75% and 70% of the species (Figure 6b, left panel) and with regions of high interannual temperature variability for 70% and 55% of the species (Figure 6b, right panel), and (c) with regions located further away from precipitation stations for 95% and 20% of the species (Figure 6c, left panel) and with regions located further away from temperature stations for 95% and 25% of the species (Figure 6c, right panel), when compared to their corresponding null distributions.

The comparisons based on the precipitation weather stations and those based on the temperature stations for interannual variability were weakly correlated (Figure 7a). However, when an outlier was not included in the regression, the correlation for false negatives was stronger ($r^2=0.74$; Figure 7a, right panel, dotted line). Correlations between precipitation and temperature for Euclidean distance uncertainty were high for both false positives ($r^2=0.99$; Figure 4b, left panel) and for false negatives ($r^2=0.97$; Figure 7b, right panel).

Among the three datasets characterizing uncertainty in interpolated climatic layers, interannual precipitation variability showed the highest values associated with false positive errors in the SDM, with a species average relative distance ($|D_{null} - D_{ov}|$) of 0.45 (where 0 represents no effect and 1 represents the highest effect). Euclidean distance to station showed the lowest species average relative distance ($|D_{null} - D_{ov}|$) of 0.26, associated with false positive errors for both temperature and precipitation based layers (Figure 8). For SDM false negative errors, the dataset that showed the highest associated value was topographic heterogeneity with a species average relative distance (

$|D_{null} - D_{un}|$) of 0.68; followed by interannual variability with a species average relative distance ($|D_{null} - D_{un}|$) of 0.47 for the temperature weather station based layer and 0.48 for the precipitation weather station based layer. Euclidean distance showed the lowest species average relative distance of ($|D_{null} - D_{un}|$) of 0.45, associated with false negatives for both temperature and precipitation based layers (Figure 8).

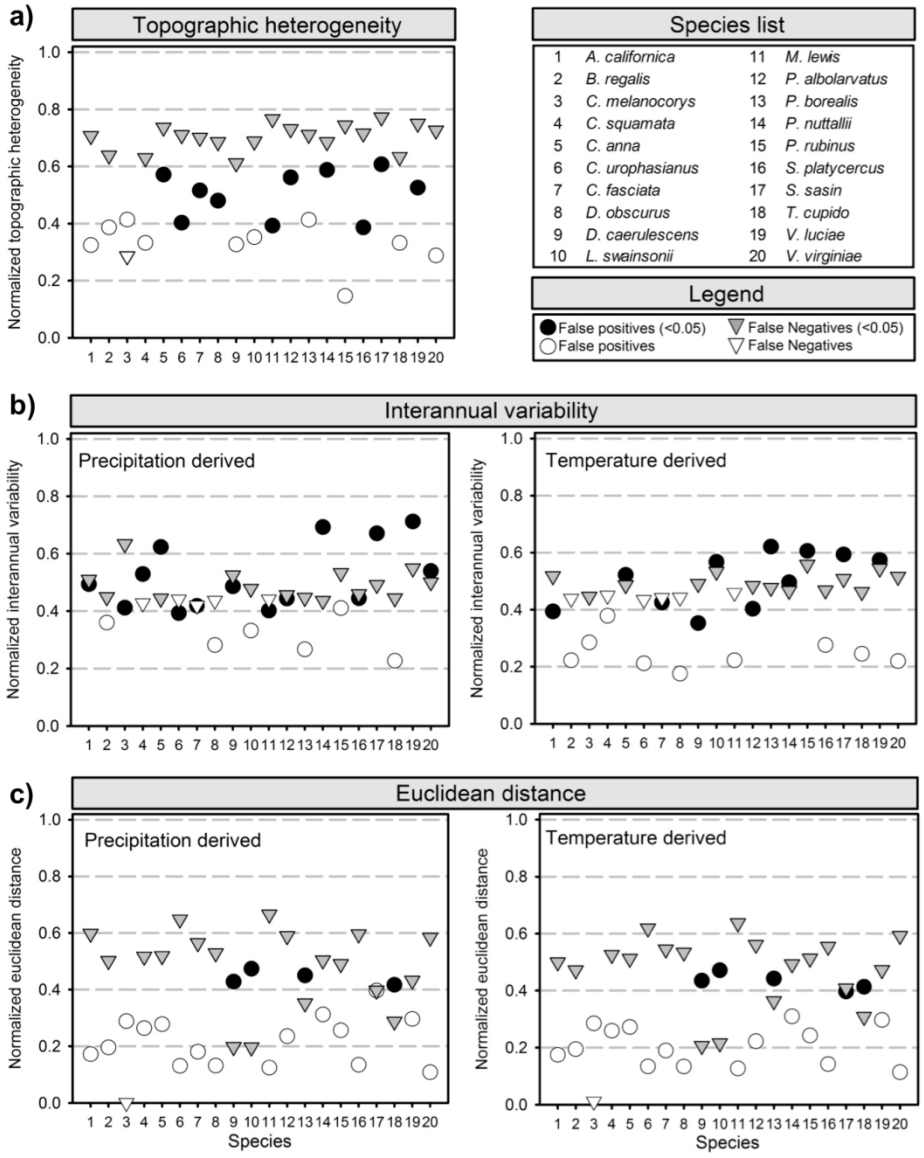

Figure 6. Species-specific normalized mean differences for each uncertainty layer. False positives are marked with circles and false negatives are marked with triangles. Statistically significant values are shown with filled markers.

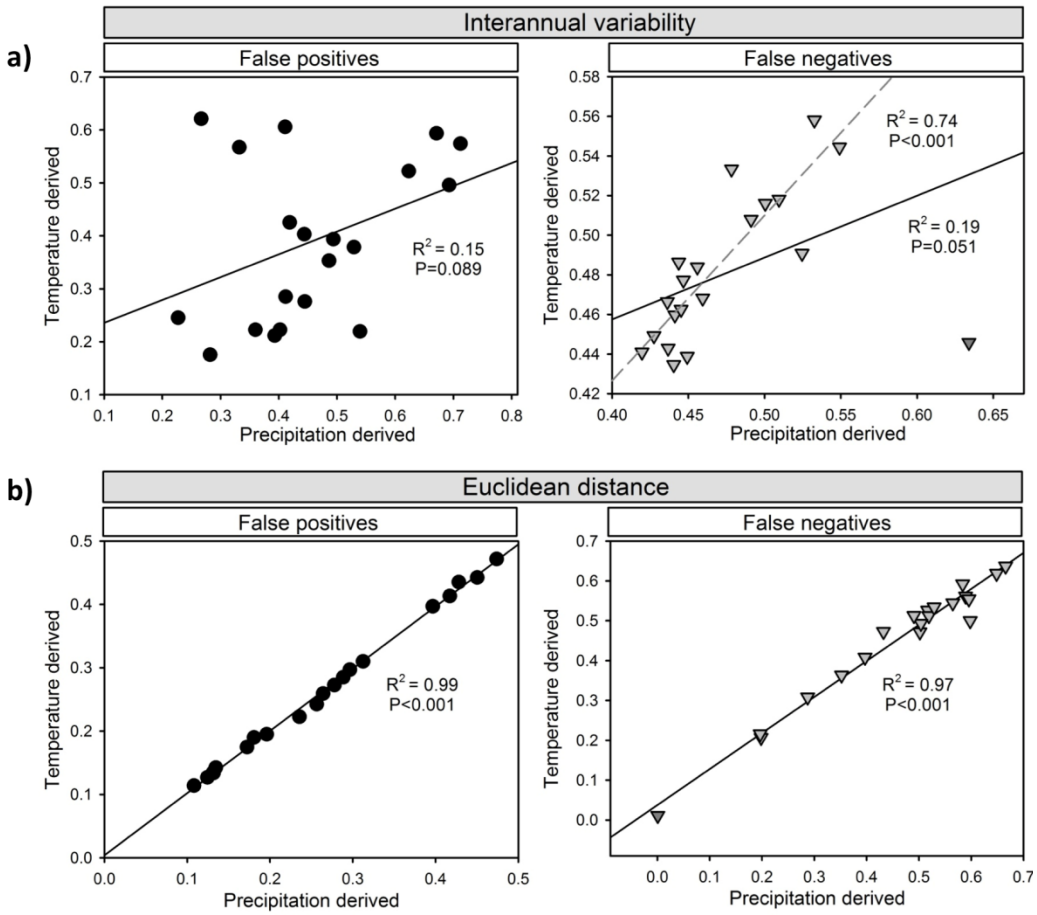

Figure 7. Comparison of temperature and precipitation results based on the species-specific normalized mean differences ($|D_{null} - D_{un}|$ and $|D_{null} - D_{ov}|$) for interannual variability and Euclidean distance. Note that for Euclidean distance the results have a high degree of similarity between temperature-derived and precipitation-derived uncertainty layers.

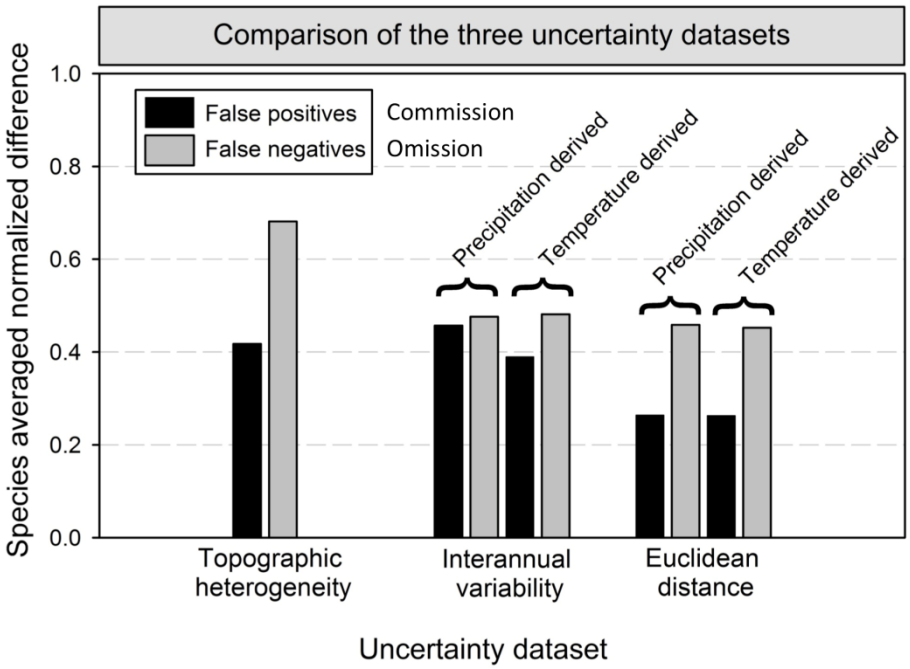

Figure 8. Results for normalized mean differences averaged across all species. Averaged values corresponding to false positives (commission errors) are shown in black and averaged values corresponding to false negatives (omission errors) are shown in grey. Note that results for interannual variability and Euclidean distance have two results, the first corresponding to precipitation and the second one corresponding to temperature.

Discussion

Our goal was to determine the effects of three different metrics of uncertainty associated with climate data geo-processing on the performance of species distribution models. Specifically, we were interested in understanding the uncertainty associated with weather station-based interpolations and its effect on omission and commission errors. We found evidence for the influence of all three of the tested sources of uncertainty on SDM performance: topographic heterogeneity, interannual variability and distance to the closest weather station.

Our results are consistent with each of our expectations; however, there are some clear differences in the degree of association between each of the uncertainty layers we analyzed and the observed errors of omission and commission. We found that, among the three uncertainty components, the highest degree of association occurred between false negative errors and topographic heterogeneity (Figure 5), indicating that in areas of high topographic heterogeneity, SDMs are more likely to under-predict than over-predict a species' distribution. This finding can be explained by how the models classify a particular pixel as species presence vs. absence. For example, if an area situated in the Great Plains in Kansas, where climatic layers have relatively low spatial variability, is compared to a second area located on the Eastern slope of the Andes where variables

change abruptly over relatively short distances, errors will diverge. In the first example the model will tend to over-predict because the climatic similarity from one pixel to the next is high and decreases slowly, making it easier for the model to fit a curve that explains the relationship between observations and the environmental layers. In the second example, the climatic similarity from one pixel to the next has the potential to change dramatically, thus making it harder for the model to fit a curve that explains the complexity in the environmental layer without raising the degree of the polynomial approximation. In this case, the model choice will be the most parsimonious solution, which results in an under-prediction. These results suggest that a model of a species that is distributed in areas of high topographic heterogeneity will have lower performance than one of a species that comes from areas of relatively low topographic heterogeneity, and that poor performance is more likely to result in under-prediction of the species' actual range.

The second largest degree of association was between interannual climatic variability and false negatives (Figure 5), which highlights the importance of establishing the appropriate temporal relationships between species observations and environmental layers as a step towards improving model performance. The processes that condition a species distribution operate at different spatial and temporal scales (e.g., Wiens 1989). Here we use interannual climatic variability, a metric that is highly synchronized with ecological processes that affect the distribution of the taxa under study. However, for some species this strategy can have implications that influence the interpretation of our results. For example, artificially influenced resource availability (e.g., bird feeders) can provide inaccurate cues regarding habitat quality (Robb et al. 2008), causing a temporal and/or spatial mismatch between the species and the natural availability of the resource. Also, natural climatic oscillations occur at multiple temporal scales, and organisms are adapted to cope with this variability at some scales better than others. For example, plants in xeric environments are well adapted to high variability in temperature at short time scales (i.e., diurnal), while a change in the pattern of interannual precipitation for the same system can have profound impacts on levels of physiological stress (Freas and Kemp 1983). To directly incorporate the effect of natural climatic variability on SDMs requires a direct connection between each particular species observation and the year when the observation was made. This is not always possible due to the lack of high-resolution gridded time series climate data for most of the world. When time-series climatic data is not available, we recommend that model confidence should be based on the life cycles of the taxa under scrutiny. For example, perennial plant species or animal species that do not migrate might be less sensitive to interannual variability than annual or migratory species.

The assumption that climatic similarity between two points in space is a function of the distance between them does not hold under every circumstance. Climatic similarity between two points will also depend on their difference in elevation; the Euclidean distance metric we used here integrates both vertical and horizontal distance to capture the effect on SDM performance of relative weather station density in topographically similar areas. Our results for this distance-based analysis show a clear difference in how

omission vs. commission errors respond (Figure 5). As an example, we can compare India, with a robust network of highly dense precipitation weather stations (0.001 stations/km^2) recording over many decades, to its neighbor Myanmar, where the density of weather stations (0.00002 stations/km^2) is two orders of magnitude less. The Euclidean distance alone is not enough to characterize the uncertainty in the interpolations of climatic parameters in both countries. It is also critical to know where these weather stations are located in relationship to the underlying environmental variability. Regions in the north of India will require a considerably larger number of weather stations to accurately characterize local precipitation patterns due to their high topographic heterogeneity. Conversely, South Myanmar may be climatically well characterized by the current density of weather stations. We advocate for the inclusion and sharing of weather station locations used to build interpolated climatic products. Open access to this information will allow users to develop their own uncertainty metrics, and to identify regions where they need to be particularly careful when interpreting the results of SDMs.

Our results are based on the assumption that the spatial distributions of the twenty bird species we analyze here are mainly determined by climate. However, for any taxon, it is unlikely that climatic variables alone will shape their realized distribution. Although examining whether climate determines these species' ranges is not the goal of this study, we acknowledge that non-climatic factors can also be responsible for a false positive or a false negative. Alternative factors responsible for absence data predicted as presence are not simple (see Lobo et al. 2010); and possibilities can be grouped into two categories: a) species related factors (e.g., locality climatically favorable but dispersal barriers prevent occurrence, interspecific interactions, local extinctions, or limited resources), and b) extrinsic factors (e.g., incomplete surveys and biased information). Because we chose to work with a multiyear observational dataset, our results should be less affected by these extrinsic factors, leaving species-related issues potentially contributing to errors. In the case of presence data predicted as absence, our analysis suggests that the underlying climatic data is not only unsuitable, but also incorrectly characterized by the interpolation due to high uncertainty. However, the underlying climate may be unsuitable and correctly characterized by the interpolation, requiring alternative explanations for why the observation was recorded as a presence. Such a result could be attributed to source-sink dynamics, transient occupancy observed by chance, or even artificial food availability. This will not change the implicit SDM assumption that documented species observations always represent suitable habitat. In other words, the MaxEnt algorithm assumes that a species will always choose the appropriate habitat. However, it is possible for an individual of a species to err in selecting climatically suitable habitat. For example, juvenile birds will have less experience in choosing locations for a nesting site; also, late arrivers will have fewer options in site selection. These outliers under normal circumstances will also have a lower chance of successfully breeding (Martin and Roper 1988). Indeed, part of the process of natural selection is individuals making mistakes in the selection of suitable habitat. These ecological mistakes, essential in the process of evolution and currently overlooked by SDM practice, have the potential to be applied to our understanding of how a species will respond to climate change.

The reasons behind our choice of taxonomic group to test our hypotheses were mainly based on species occurrence data quality and availability, which also limits inference and application to other taxonomic groups. Yet we believe that the same principles and mechanisms explored here apply also to other organisms, and that the results can be cautiously extrapolated to other taxonomic groups. However, underlying quality and accuracy of the gridded information used as environmental layers should not be the only direction to look for sources of over- and under-prediction errors and ways to improve the models. Coastal redwood (*Sequoia sempervirens*), as an example of a relict plant species, has its suitable bioclimatic envelope restricted to a narrow 50 km belt in the coast of California. However, some parts of its actual current distribution may be better explained by factors of land use change rather than climate (Pyke 2004).

Attributing the omission and commission errors in areas that were predicted by the model as presence due to climate alone will be incorrect. A simple yet difficult-to-achieve recommendation for resource managers applying the results of SDMs is to asses all the potential sources of uncertainty and to focus on characterization of the sources that provide the largest amount of error. However, integrated tools to aid resource managers in evaluating multiple sources uncertainty, while needed, are not yet available.

Finally, a useful SDM is not only precise, but also accurate. While past literature that deals with SDM environmental layer uncertainty focuses on model precision, (e.g., Kriticos and Leriche 2010) here we ask how divergent results can be between known and modeled distributions if a parameter is uncertain. We quantified the relationship between omission and commission errors in the predictions and the degree of uncertainty in the interpolated environmental input layers. We attribute decrease in SDM performance to the three aspects of uncertainty evaluated here; however, not all of them were identified as equally important sources of over- and under-prediction errors in SDMs. Our results confirm the importance of establishing appropriate relationships in time and space between species and environmental layers. Uncertainty characterizations for environmental layers can provide operational criteria for the selection of species observations fed into SDMs, and help identify conditions where users can weigh their degree of confidence when making decisions based on a SDM.

CHAPTER 3
Back to the future: using historical climate variation to project near-term shifts in habitat suitable for coastal redwood

Introduction

Although global climate models (GCMs) project changes in climatic patterns at coarse spatial scales, the regional to local manifestations of climate change are not yet well quantified (Knutti and Sedlácek, 2013), limiting our ability to incorporate them into ecological forecasts (Osmond et al., 2004). Further, local climatic trends do not always follow global trends (Helmuth et al., 2002), suggesting a considerable need to identify how global forcings are intensified or weakened by local conditions (Cordero et al., 2011). One example of the inability of GCMs to capture complex local climate in future simulations occurs in California and other western continental margins, where local conditions are characterized by a cool coastal climate and substantially warmer interior. While weather station data in California show a coherent statewide positive trend in minimum surface air temperature (LaDochy et al., 2007), maximum temperature trends vary spatially, with cooling in coastal areas and warming in inland areas (Lebassi et al., 2009), suggesting that the local manifestation of large scale warming is affected by local negative feedback processes along the coast.

The mechanism proposed to explain the asymmetric change in surface air temperatures involves differential heating between ocean and the coastal landmasses, which has been hypothesized to result in stronger and more persistent wind-driven coastal upwelling along the coast of California (Bakun, 1990). Wind-driven coastal upwelling is the product of wind stress moving parallel to the shoreline toward the equator that, combined with the earth's rotation, displaces surface water offshore and brings pulses of deep cold water from beneath (Mann and Lazier 2006). Based on regional climate models, intensification of wind-driven coastal upwelling has been postulated to limit the projected increases in temperatures by reducing insolation and raising humidity in coastal terrestrial ecosystems in California (Snyder et al., 2003; O'Brien et al., 2012). However, although long-term observations support this prediction (Seo et al., 2012), it has not yet been consistently corroborated by GCM simulations (Wang et al., 2009), likely due to oversimplifications in coupled ocean-atmosphere models and the mismatch in spatial scale between the coastal intensification and resolutions of global simulations (Bakun et al., 2010).

The limited capacity of global simulations to resolve local climates has profound implications for projecting climate change impacts in many fields (Wilby et al., 2004). Species distribution modeling (SDM) is one of these areas (Kremen et al., 2008). SDM integrates taxonomic and geographic data associated with documented species occurrences and interpolated climatic observations to produce a set of correlative rules that identify the multidimensional space where the species was collected or observed (Peterson and Vieglais, 2001). This n-dimensional space can then be projected back onto the same geographic space that was used to generate it, yielding a very fine scale

hypothesis of current climatically suitable habitat. SDMs can also be applied to estimate where the climatic conditions will be suitable for the species in the future under the assumption that the species-climate relationship remains stable (Franklin, 2009). This is achieved by projecting fine scale hypotheses of current suitable habitat into downscaled future global simulations. Downscaling is a necessary process that brings down the spatial resolution of the coarse scale climate models into the resolution of current climate datasets to avoid combining data with different discretization levels (Wilby et al., 2004).

Multiple approaches are used to downscaled global simulations; they range from basic approaches such as the simple change factor approach to more complex techniques such as statistical and dynamic downscaling. The simple change factor approach takes into consideration the change in a particular climatic parameter from one time period to another, and imposes the resulting difference to observed climate in order to obtain the future downscaled data (e.g., Tabor and Williams, 2010). The problem with this approach is that it is unable to get higher order climate statistics that are sometimes required in vulnerability assessments, and also assumes homogenous climatic anomalies. In the intermediate category is the synthetic statistical downscaling approach (e.g., Wood et al., 2004). This approach is able to incorporate measures of variability and it doesn't need to be bounded to a specific region or few models, which makes it widely employed for generating multiple future scenarios. Unfortunately, some of the products based on synthetic statistical downscaling typically inherit the limitations of simulations at coarse spatial scales and are often unable to resolve local manifestations of climate. A second limitation of some of the products from this approach is that climate variables are downscaled independently, sometimes resulting in physically implausible outcomes. Dynamic downscaling (e.g., Castro et al., 2005), is a more sophisticated approach that can reproduce local atmospheric processes with relatively high fidelity, but is limited to small regions and few models due to the computational cost. Also, the products of dynamical downscaling, known as regional climate models (RCMs), typically do not represent coastal processes and, therefore, would not include dynamic response of upwelling to large-scale forcing. Finally, under the name of deterministic statistical downscaling (e.g., Hidalgo et al., 2008; Abatzoglou and Brown 2012), there is a group of methods that try to address some of the limitations of synthetic statistical and dynamic downscaling approaches. The deterministic statistical approach is computationally more efficient than the dynamic statistical approach and it doesn't need to be bounded to a particular region. Moreover, it overcomes the uncertainties associated with spatial interpolation that exist in the synthetic statistical approach. This method benefits from the use of historic climatic data that already captures regional climate and weather patterns with high fidelity; however, at the same time, it is limited by the spatial resolution of the data used to produce it, since it requires relatively high frequency historic climatic data to construct the analogs (Abatzouglou and Brown, 2012). A second limitation of this approach is that it is currently restricted to the few GCMs that archive daily data.

Researchers that attempt to project the potential effects of climate change on biodiversity using SDMs and downscaled climatic outputs should base the decision about what gridded climatic dataset they will use on the methods and limitations from each

downscaling approach and the specific questions they want to answer. However, this is not always possible and, often, practical considerations have a higher relative weight than methodological considerations in the final decision. If the spatial resolution is the most important criteria for the selection of the appropriate dataset, probably the datasets that result from the basic downscaled approach (e.g., Tabor and Williams, 2010) should be used; however, these products can't resolve local manifestations of climate such as the effect of wind-driven coastal upwelling on surface air temperature and humidity patterns on coastal terrestrial ecosystems. If the skill of capturing the sharp energy/moistures gradients between coastal and inland essential to coastal terrestrial ecosystems is the most important criteria for the selection of the most appropriate dataset, probably the one that results from deterministic statistical downscaling approach should be selected.

Currently, there isn't an available downscaled dataset produced in such a way that can capture both the effect of wind driven coastal upwelling on surface air temperature and humidity patterns on coastal terrestrial ecosystems and, at the same time, present a spatial resolution fine enough to capture the high environmental heterogeneity characteristic of costal California; a combination of properties that would allow us to delineate conservation strategies at a scale relevant to resource management decisions. Despite this issue, researchers have used dynamically downscaled (Kueppers et al., 2005) and statistically downscaled temperature and precipitation outputs (Loarie et al., 2008; Klausmeyer and Shaw, 2009) with SDMs to project potential effects of climate change on California biodiversity. These SDMs do not include the already observed changes in wind-driven coastal upwelling, nor the predicted effects on surface air temperature and humidity patterns on coastal terrestrial ecosystems, which suggest that the range shifts estimates for species that live within regions influenced by ocean-atmosphere interactions may not reflect realistic changes in temperature and precipitation across their distribution.

Projecting biodiversity response to climate change is a crucial step towards mitigation and adaptation strategies. As a consequence, we cannot afford to wait until global simulations are capable to capture fine scale atmosphere-ocean interactions that exist in climatologically complex and biologically unique regions as coastal California (Loarie et al., 2008). The aim of this study was to develop a novel approach to more reasonable estimates of climatically suitable habitat that incorporates not only the regional but also the local manifestation of greenhouse induced climate change for a coastal species, which distribution coincides with a region that is strongly influenced by ocean-atmosphere interactions. Coastal redwoods (*Sequoia serpervirens*), once a widely distributed species (Barron et al., 2003), and now limited to a narrow 50 km belt along the coast of California (Noss, 2000), is particularly well suited for our analysis. Redwoods also provide an interesting case study because they are known to be poor regulators of water usage (Burgess and Dawson 2004). They transpire considerable quantities of water at night because they are unable to fully close their stomata (Burgess and Dawson 2004). However, *in situ* measurements have shown that the presence of cloud moisture in the canopy can nearly eliminate the water vapor deficit that causes transpiration (Simonin *et al.*, 2009). These physiological traits and restricted distribution are evidence that redwoods rely on high levels of ambient humidity, suggesting that SDM based on

temperature and precipitation alone may be inadequate.

In this paper we address four questions: (1) Does adding monthly water deficit, a variable directly related to redwood physiology, improve redwood distribution model accuracy? (2) Do projections of future mean regional climate have observed analogs in the historic climatic record, for example as individual years in the tails of annual temperature distributions? (3) Do climate scenarios that more reasonably capture local manifestations of climate change differ from those obtained from coarse resolution global climate models? (4) Do scenarios of climatically suitable habitat that integrate historic climate variability allow us to identify highly stable zones (refugia) for conservation planning and management in coastal California? To address these questions, we evaluated species distribution models with and without monthly water deficit, in addition to the more typical temperature and precipitation variables. We then used naturally occurring variability in the historic climatic record to develop multiple scenarios of California climate and compared these to GCM projections from the CMIP5 archive. We developed estimates of climatically suitable habitat for redwoods under historical "normal" temperature and precipitation, and for the multiple climate scenarios. With this approach, we maintain coherent relationships between regional climate and local effects of coastal upwelling, and also preserve observed covariance among climatic variables in space and time. Combining the scenarios we developed into an ensemble model allow us to identify coherent sub-regions robust to modest climate change that can inform management and conservation initiatives in coastal ecosystems in California.

Methods

We used the tails of the normal distribution of temperature and precipitation values (Hansen et al., 2012), to develop scenarios of California coastal climate at finer spatial scales than those produced by global climate models (GCMs). This new approach, similar to the climatic-analogs method developed by Lorenz (1969), takes advantage of spatially gridded climatic time series. However, instead of searching for equivalent surfaces at the resolution of the GCMs (Zorita and Von Storch, 1999), we first selected climatically anomalous years at the statewide scale from the historic record and matched them to near-term changes predicted by GCMs to limit the range of plausible scenarios and provide a temporal context.

Selection of anomalous years in California

An anomalous year is defined here as an unusually warm or cool year for temperature (wet or dry years for precipitation) in the historical climatic record. They are identified as years in which the annual departure, relative to a baseline, falls in the tails of the normal (Gaussian) distribution (i.e., values are above or below one standard deviation from the mean; Hansen et al., 2012). To identify historic anomalous years in California, we used a dataset hosted at the California Climate Data Archive (CCDA; Abatzoglou et al., 2009). This dataset was produced by the Western Regional Climate Center and Scripps Institution of Oceanography and includes a network of 195 Cooperative Observer

Network climate stations for the State of California. This dataset avoids potential bias arising from fluctuations in spatial and temporal coverage of weather stations by filling missing values using data from an interpolated gridded time series (PRISM; Daly et al., 2000). We did not use PRISM directly to avoid, as much as possible, any uncertainties associated with the interpolation (Fernandez et al., 2013). From the CCDA dataset, we examined four annual climatic variables, but here we only focus on mean annual temperature and annual total precipitation; results for maximum and minimum temperature are included in the Supplementary information (*Appendix C*, Figs. S45-S48).

We computed departures by subtracting the observed mean annual temperature and annual total precipitation values for each year from statewide averages for the full historic record (i.e., 1895 to 2010) (Fig. 9). The departures for each year (x) and the values that represent +/- one standard deviation (s) from the mean (μ) were used to assign each year to one of three possible categories: (1) $x > \mu + s$; (2) $x < \mu - s$; and (3) $\mu - s < x < \mu + s$. We classified years based on both annual mean temperature and total annual precipitation, thereby, assigning each year to one of nine quadrants (Fig. 10). The years within each of the eight periphery quadrants represent rare (i.e., less probable) climates, and we used them as climatic scenarios in our analysis (Fig. 10). The years represented in the central quadrant are more common (i.e., more probable) combinations of temperature and precipitation, and we use them to characterize “normal” climatic conditions.

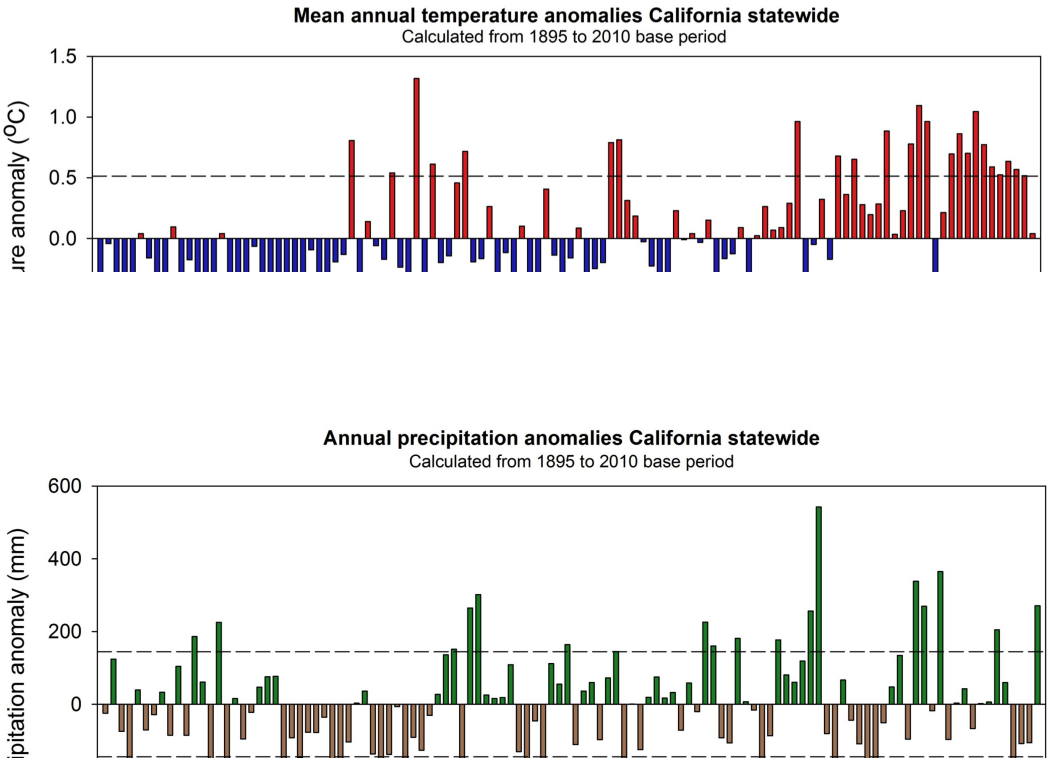

Figure 9. Mean annual temperature anomalies from 1895 to 2010 baseline for California (top) and annual precipitation anomalies (bottom) based on the CCDA dataset. Dotted line represents +/- one standard deviation from the 1895-2010 mean (solid line) for each variable.

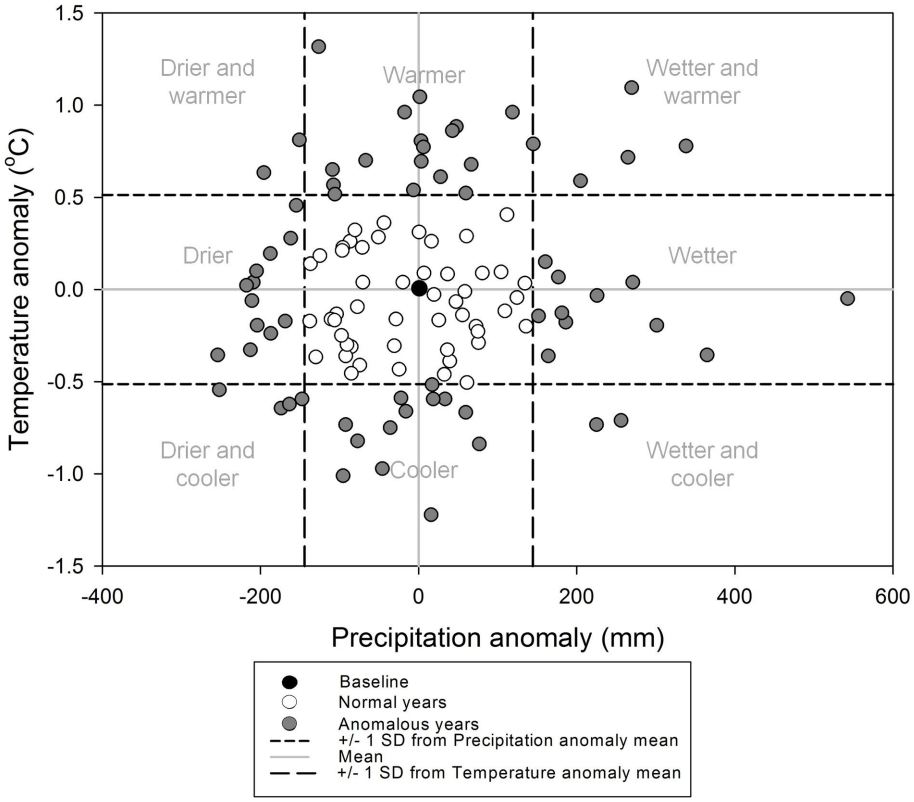

Figure 10. Assignment of individual years to climate scenarios. Each white or gray circle represents mean annual temperature and total precipitation anomalies for an individual year between1895 and 2010. The black circle in the center represents climatologically "normal" conditions obtained by calculating means for the full period. Dotted lines represent +/- one standard deviation from the 1895-2010 mean (grey line) for each variable (short dotted line for temperature and long dotted line for precipitation).We use the years grouped in the eight periphery quadrants as the multiple climatic scenarios in our analysis.

Global climate change context

The above climatic scenarios reflect naturally occurring climatic variability that redwoods have experienced in the past century. To relate these scenarios to plausible future conditions and to provide a temporal context for our analysis, we compared these historically-based scenarios to GCM output from the Coupled Model Intercomparison Project (CMIP5) Representative Concentration Pathway (RCP) 4.5 (Thompson et al., 2011). The RCP4.5 greenhouse gas concentration trajectory provides a conservative estimate of global temperature change for the future compared with alternative trajectories (Peters et al., 2013), although trajectories of temperature change do not diverge substantially until after the 2040's (Knutti and Sedlacek, 2013). From the CMIP5 archive, we obtained mean monthly temperature and total monthly precipitation from 19 GCMs for the period 1895 to 2099. We calculated mean annual temperature, annual total precipitation and anomalies for each year from 2020 to 2099 using the same baseline

period used in the selection of historically anomalous years (i.e., 1895 to 2010) for each GCM. We compared these future model-derived anomalies to the anomalies obtained for each year in the observed historical period (*Appendix C*, Figs. S57-S60). We then calculated mean anomalies for three future 20-year periods (i.e., 2020 to 2039, 2040 to 2059 and 2060 to 2079).

Climate variables for species distribution models

To generate the SDMs we used a gridded time series dataset for the period of 1895 to 2010, at a resolution of 800 m, from the PRISM Climate Group at Oregon State University (Daly et al., 2000). For every grid cell, PRISM values are estimated using a local regression where surrounding weather stations used to populate the regression are weighted by their physiographic similarity to the grid cell being modeled. PRISM incorporates the effect of elevation, terrain-induced airmass blockage, coastal effects, temperature inversions and cold-air pooling; and includes monthly gridded layers for dew point, precipitation, temperature and vapor pressure, providing one of the finest spatial resolutions relative to other similar products at regional scale (but see Flint and Flint 2012). PRISM also constitutes a well vetted and critical resource for studies in diverse scientific fields including ecology, biogeography, conservation and natural resource management (e.g., Fitzgerald and Gordon et al., 2012; Franklin et al., 2013; Torregrosa et al., 2013), which focus within the continental U.S.

In addition to PRISM monthly variables (i.e., maximum, minimum temperature, total precipitation, mean dew point temperature and vapor pressure), we decided to include in our analysis a variable that addresses the eco-physiological challenges that redwoods face. Previous research suggests that the seasonal interactions between energy and water supply could potentially improve our understanding of the role of climate in defining vegetation distribution at fine spatial scales (Stephenson, 1990). From climate water balance calculations (Stephenson, 1998), we could derive climatic water deficit (*WD*) as the difference between potential and actual evapotranspiration. However, actual evapotranspiration is hard to accurately estimate, as it is a function of the amount of soil moisture, soil type, vegetation type and slope. Moreover, the two sources of soil information in the U.S.: the Soil Survey Geographic Database (SSURGO) and the U.S. General Soil Map (STATSGO2) don't have complete coverage for the area of study. To solve this problem, we followed Stephenson (1998) recommendation and thus decided to use a different approach to calculate the seasonal interaction between energy and water supply (Ellis et al., 2008). We calculated *WD* as the difference between precipitation and potential evapotranspiration (Paltineanu et al., 2009; Ellis et al., 2010). We calculated water deficit for each month of every year for the period of 1895 to 2010 for the continental U.S. using the Hamon (1963) method, which is a refinement of the Thornthwaite method (Thornthwaite and Matter, 1955). This method has been proven to be robust under a wide range of conditions (Vörösmarty et al., 1998) and, when compared to alternative approaches, it provided the most accurate approximation when the inputs are limited to temperature and precipitation (Lu et al., 2005).

Monthly water deficit (*WD*) was calculated as follows:

$$WD = P - PE$$

where *P* represents precipitation for each month, and *PE* represents monthly potential evapotranspiration. *PE* is calculated as:

$$PE = 13.97 d D^2 W_t$$

where *d* represents the number of days in a month for a particular year, *D* represents mean monthly daylight hours in units of 12 hours, and W_t represents saturated water vapor density. W_t is calculated as follows:

$$W_t = \frac{4.95 e^{0.062T}}{100}$$

where *T* represents mean monthly temperature in (^{o}C). *D* is calculated following Forsythe et al. (1995):

$$D = 24 - \frac{24}{\pi} \cos^{-1} \left(\frac{\sin \frac{0.8333\pi}{180} + \sin \frac{L\pi}{180} \sin \phi}{\cos \frac{L\pi}{180} \cos \phi} \right)$$

where *L* represents the latitude and ϕ represents the sun's declination angle. ϕ is calculated as:

$$\phi = \sin^{-1}(0.39795 \cos \theta)$$

where θ represents the revolution angle. θ is calculated as:

$$\theta = 0.2163108 + 2 \tan^{-1} \left(0.9671396 \tan \left[0.0086 \left(J - 186 \right) \right] \right)$$

where *J* represents the day of the year.

Based on the PRISM variables and the derived monthly *WD*, we created two configurations of climate layers: (1) PRISM; (2) PRISM+*WD*.

Species occurrence data

To build the species distribution models, we compiled all of the *S. sempervirens* occurrence data from museum specimens, primarily from georeferenced specimens from the Consortium of California Herbaria (accessed June 2012), a centralized repository for 16 regional herbaria, and from the Global Biodiversity Information Facility (accessed June 2012), a global consortium of biodiversity data holding institutions. To evaluate the models, we obtained independent occurrence data from redwood specimens held at the California Academy of Sciences, which were retrospectively georeferenced following Chapman and Wieczorek (2006) protocols. In this study, a species occurrence was defined by a unique locality expressed as latitude and longitude with positional uncertainty represented by a maximum error estimate of less than 800 m, supported by a vouchered specimen collected from 1895 to 2010.

Species distribution models

We generated species distribution models using MaxEnt v3.3.3k (Phillips et al., 2006), a method particularly effective at dealing with presence-only data (Elith and Leathwick 2009). MaxEnt uses the principle of maximum entropy to estimate a set of rules correlating environmental variables and species occurrences to approximate the potential bioclimatic habitat of the target species (Phillips and Dudik 2008). The MaxEnt algorithm is related to Bayesian theory and considers redundant information without over fitting; eliminating the need to apply a variable reduction technique before running the models (but see Parolo et al., 2008). We calibrated the models using an approach that addresses spatial autocorrelation by using a spatially structured partitioning procedure adapted from Fernandez et al., (2013). This process randomly resamples species' observations into different subsets where 80% of the localities are used for training and 20% are used for testing the model. We repeated the process creating 100 subsets of the species observations that we used in turn to produce 100 MaxEnt models for "normal" conditions (i.e., averages of variables across the years within the central quadrant; *Appendix C*, Figs. S49-S54) based on the two environmental layer configurations (i.e., PRISM and PRISM+*WD*). We used the default values in the MaxEnt algorithms for the maximum number of iterations and convergence threshold (i.e., 500, 10^{-5}).

We averaged the 100 bootstrapped results based on the PRISM and the PRISM+*WD* layers configurations into two final niche models that provided a continuous index of relative suitability under "normal" conditions with and without *WD*. We converted these probabilistic outputs into a presence/absence map based on a widely accepted method of using the value of the points on the receiver operating characteristic curve where the sum of sensitivity and specificity is maximized (Loarie et al., 2008).

We evaluated the final models produced for "normal" conditions under the two environmental layer configurations using several complementary approaches. First, we took advantage of the independent evaluation data and measured the prediction success of the models for the withheld independent localities classified as presence (Zweig and Campbell, 1993). Since prediction success is a function of the threshold for assigning presence, we also evaluated the models using the area under the receiver operating characteristic curve (AUC). AUC is a threshold-independent metric that summarizes a model's overall performance over every possible threshold (Lobo et al., 2008) and although it has been criticized for being sensitive to the total extent over which models are produced (VanDerWal et al., 2009), this should not affect our results because we kept extent constant. Although these metrics are statistically defensible, they are still hampered by the lack of true absence data (Lobo et al., 2010). Since the current distribution of redwoods is relatively well known and mapped, we compared the models for "normal" conditions created with and without *WD* to the known distribution of redwoods using the true skill statistic (TSS; Allouche et al., 2006). We used the Classification and Assessment with Landsat of Visible Ecological Groupings (CALVEG 2009), a fine scale dataset that was primarily derived from remote sensing and produced by the U.S. Department of Agriculture, as the known distribution.

We projected each bootstrapped simulation that contributed to the final model for the "normal" conditions based on the environmental layer configuration that produced the best results (i.e., PRISM+*WD*) into each of the anomalous years. Since each scenario has a different number of years we combined the probabilistic outputs from MaxEnt for each group of years before selecting the threshold following Marmion et al., (2009). We converted the results into presence/absence maps using the same threshold rule used for "normal" conditions.

Analysis of the distributions

To quantify changes in the distribution of suitable habitat, we compared the projected distributions for individual anomalous years to those for "normal" conditions based on the environmental layer configuration that included *WD* using three complementary approaches. First, we evaluated the differences in spatial patterns between current and projected scenarios by subtracting gridded model outputs under each scenario from the distribution under "normal" conditions and mapped the differences. Second, we measured the change in area between the current distribution and the distributions under each of the eight possible climate scenarios. To calculate areas, we converted all gridded model outputs into polygons and projected them into an equal area projection (Albers Equal Area Conic); we measured areas using Spatial Statistics toolset in ArcGIS v.10.1. Third, using centroids we measured the change in distance and direction between the current distribution and the distributions under each of the eight possible climate scenarios. To calculate centroids we projected the data into Azimuthal Equidistant projection and measured the shifts in distance and direction relative to the "normal" distribution using the centers of mass using the Geographic Distributions toolset in ArcGIS v.10.1.

Results

Models for the "normal" conditions based on both PRISM and PRISM+*WD* environmental layer configurations yielded reasonable approximations of the known current distribution (Fig. 11). When both SDMs were compared using the independent species occurrences for model validation we found that the prediction success for the PRISM and PRISM+*WD* environmental layer configurations were 96% and 94.5% and the value for AUC were 0.931 and 0.952 respectively (where a value of 0.5 indicates a model that is not better than random and a value of 1 indicates a theoretically perfect model). When the known current distribution from CALVEG was compared to the SDMs for the "normal" conditions, the true skill statistic was 0.956 for the SDM based on PRISM and 0.988 for the SDM based on PRISM+*WD*.

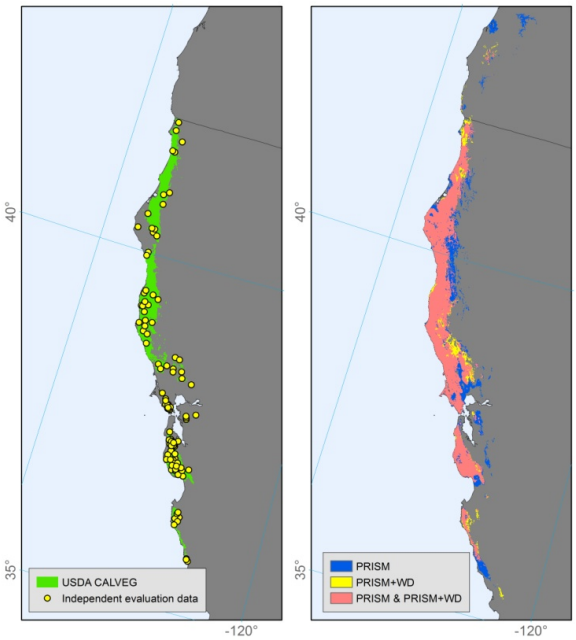

Figure 11. Model evaluation. Left, model evaluation using independent occurrence data obtained and georeferenced from the California Academy of Sciences. Right, model evaluation using a remote sensing derived dataset (USDA CALVEG).

Our results showed that the historical climate scenario that best resembles the future climate scenarios is the one where the temperature is higher than average but precipitation remains relatively unchanged statewide (Fig. 12). Moreover, when we compared model anomalies for multiple years, we found that most of the years for the warmer (normal precipitation) scenario that we developed are within one standard deviation of the 2020s and 2030s from the future but also that there is a trend in future models that show that temperature keeps increasing but precipitation remains within the normal range (Fig. 12). Although the years from the future models are centered in the warmer (normal precipitation) scenario, the standard deviation across models for the annual total precipitation extends into the drier and warmer scenario as well as the wetter and warmer scenario.

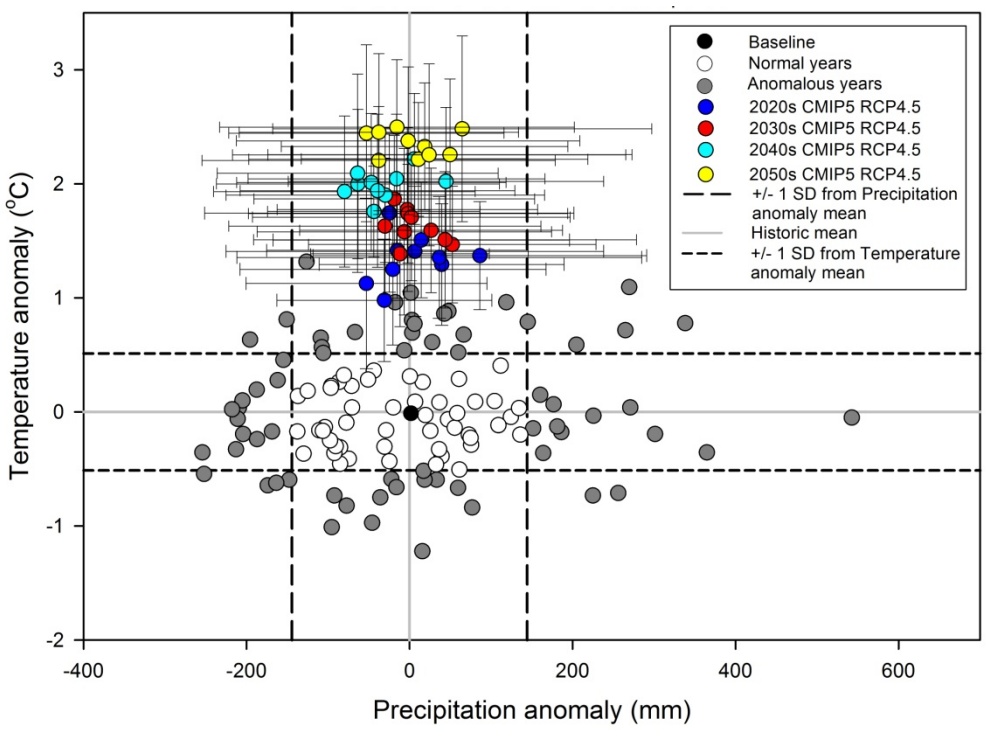

Figure 12. Multi-model mean annual temperature and precipitation anomalies for California projected for three decades in the 21st century (colored circles) compared to historical annual temperature and precipitation anomalies (circles as in Figure 10).

The potential distribution under each scenario (*Appendix C*, Figs. S49-S54) was synthesized using a map that integrates the results across years in each category and yields an ensemble estimate of climate change induced distributional expansion/contraction for redwoods for each scenario (Fig. 13). The distribution of redwoods projected into the scenario that resembles closely the CMIP5 projections of regional mean climate showed a range contraction in the south, with a reduction equal to 8.809 km^2 (50% of the climatically suitable area under "normal" conditions; Fig. 14), with no suitable habitat remaining south of the San Francisco Bay. This contraction is balanced by an expansion in the north, with a gain in area equal to 5,895 km^2 (34%; Fig. 14). The stable area for this scenario is restricted to the coastal areas in the central part of the distribution (Fig. 13B).

In addition all the scenarios showed some degree of stability, contraction and expansion in projected geographic ranges. The three drier scenarios (Fig. 13A, 13D and 13F) produced greater contraction and less expansion compared to the remaining five scenarios (Fig. 14). Similarly, the three wetter scenarios (Fig. 13C, 13E and 13H) showed the least contraction relative to the remaining scenarios, with slightly less contraction as scenario temperatures cooled (Fig. 14,); these three scenarios also had the greatest expansion. The three warmer scenarios (Fig. 13A, 13B and 13C) and the three cooler scenarios (Fig. 13F, 13G and 13H), both showed high and low degrees of contraction

within each group, and then note that for a given precipitation change (Fig. 14,), how temperature affects the distribution (Figs. 13ABC and 13FGH).

When the individual scenarios were inspected, the drier and warmer scenario (Fig. 13A) showed the highest degree of contraction, with a reduction equal to 13,691 km^2 (79% of the climatically suitable area under "normal" conditions; Fig. 14). Most of the stable habitat under this scenario was located north of 40^o latitude (Fig. 13A). This same scenario projected an increase in area of 847 km^2 (5%), mainly north of the California-Oregon border (Fig. 13A). The drier scenario (normal temperatures; Fig. 13D) presented the next highest degree of contraction with a loss in suitable habitat of 10,879 km^2 (63%); Fig. 14), closely followed by the drier and cooler scenario (Fig. 13F) with a loss in suitable habitat of 10,610 km^2 (61%; Fig. 14). Most of the stable suitable habitat was restricted to coastal areas in both scenarios (Fig. 13D and 13F).

The wetter and cooler scenario (Fig. 13H) had the highest degree of expansion with a significant estimated gain equal to 7,761 km^2 (44% of the climatically suitable area under "normal conditions"; Fig. 14), mostly located in the coastal areas at the north and south extremes of the distribution; and away from the coast in the central part of the distribution (Fig. 13H). While this scenario resulted in the largest relative expansion, it also projected some reduction discretely localized north of the San Francisco Bay (Fig. 13H). The second largest degree of expansion was presented by the wetter and warmer scenario (Fig. 13C) with a projected increase of 6,380 km^2 (37%; Fig. 14). The scenario that presented the lowest degree of habitat expansion was the drier and warmer (Fig. 13H) with an estimated gain of 847 km^2 (11%; Fig. 14).

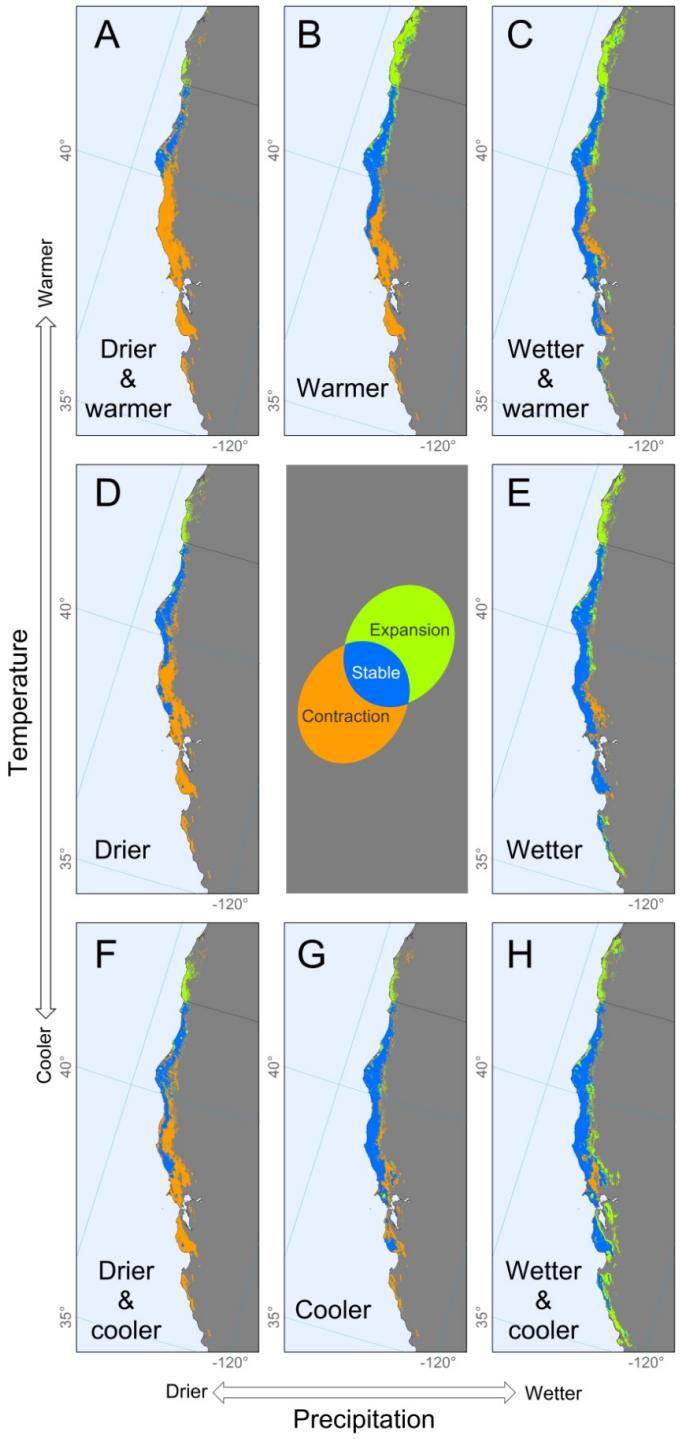

Figure 13. Synthetic generalization of the predicted expansion, contraction and stability for the eight scenarios we developed.

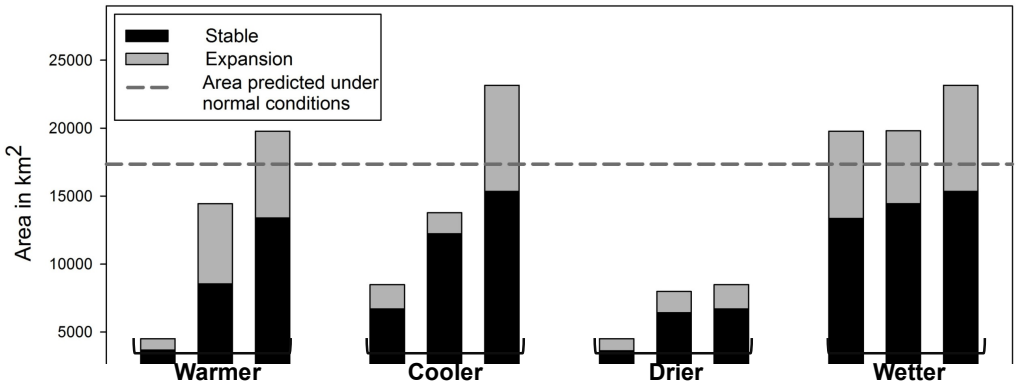

Figure 14. Comparison of changes in absolute area among scenarios.

When the centroids of each projected distribution were compared to the centroid of the historical mean redwood distribution, the scenarios with the greatest relative shifts were the warmer scenario (normal precipitation; Fig. 13B) and the drier and warmer scenario (Fig. 13A), with shifts of 201 and 192 km, respectively (Fig.15). The wetter and cooler scenario (Fig. 13H) showed the lowest degree of geographic shift of 10 km (Fig. 15). All the scenarios, except the wetter and cooler scenario presented an average north-northwest (insert mean +/-5^{o}) direction of displacement (Fig. 15).

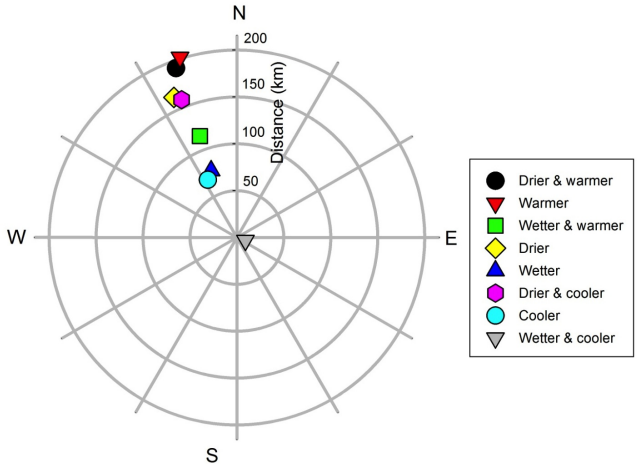

Figure 15. The shift in the center of mass of modeled redwood distributions for each of the eight scenarios relative to that under "normal" conditions.

When we combined the stable areas from all the scenarios that integrate historic climatic variability to characterize coherent sub-regions robust to modest climate change we

found that the most stable region is located in the northern part of California (Fig. 16) restricted to an area of 3,010 km^2. When we combed only the stable areas from only the warmer scenarios the area was of 3,642 km^2.

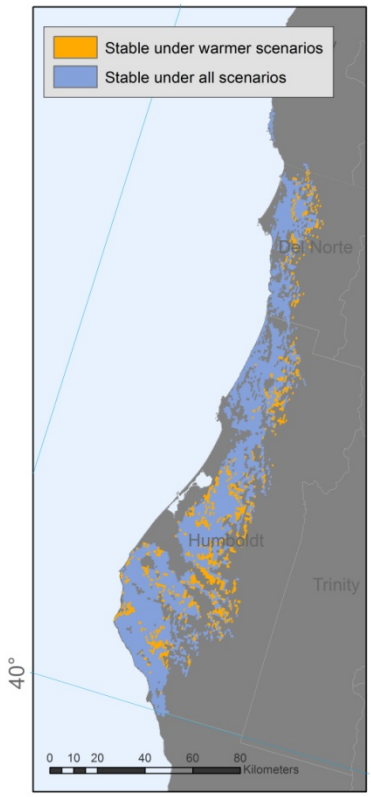

Figure 16. Ensemble scenarios for climatically stable sub-regions.

Discussion

Our first goal was to improve model accuracy by developing a tailored variable customized to redwood physiology. Different from what we expected, we did not find statistically significant improvement in the performance metrics between the models based on PRISM alone and PRISM+*WD*, although we did find small differences. It is clear that prediction success on independent data alone is not a very informative metric; the environmental layer configuration that did include *WD* was rated slightly below the one that didn't include *WD*. However, acknowledging that this metric is highly influenced by the area predicted as presence, it is possible to get a very high prediction success value with a model that tends to over predict. On the other hand, AUC and TSS ranked the models in similar way, giving the model based on the PRISM+*WD* a slight edge over the model based on PRISM alone, even though we use pseudo-absence data to calculate the AUC values, and the TSS was based on a cell by cell comparison to the CALVEG dataset. Overall it was informative to know that the prediction success on independent data was high for both models. However that the AUC and TSS metrics converge in their results was more helpful in selecting the best performing model because

they were calculated using different datasets and both account for how much area is over predicted, therefore they are more sensitive to the errors of commission in the SDM. In addition to the value provided by performance metrics, an important aspect of model evaluation is the subjective assessment that users apply to models. While rating models based only on a single number that attempts to capture the overall model performance in the landscape can result in no apparent improvement, once these performance metrics are translated into the geography, small differences can become very important. Moreover practical applications of species distribution models in conservation planning that are focus on a particular regions (e.g., reserve design and species invasions) do not value all areas in the map similarly and the overall improvement of the model is often less important than the capacity of the model to correctly predict small areas that are under scrutiny or for which prior knowledge of the species distribution is available.

The slight improvement we obtained in the model based on PRISM+*WD* over that based on PRISM is telling us that we are looking in the right direction; but that we couldn't obtain conclusive evidence that supports the use of *WD* to improve the model suggest that either the choice of variable or the method we use is not appropriate. *WD* can be calculated in multiple ways that can include additional variables such as soil depth, slope orientation and wind speed (Lu et al., 2005). However in our experiment we were limited to temperature and precipitation and the only additional information besides these two variables that we included was a calculation of daylight hours that provides a measure of the amount of energy received corrected by latitude and time of the year and assumes a topographically homogeneous landscape. We can't discard a third option that results from the fact that both models, with and without *WD,* were able to capture relatively well the known distribution of coastal redwoods. This suggests that perhaps redwood physiology was already addressed by a variable included in the PRISM original dataset and the inclusion of *WD* in the SDMs did not add any additional information that was not already there. We suspect of monthly vapor pressure which is part of PRISM, however we did not run any comprehensive test to test this hypothesis.

We took advantage of naturally occurring climatic variability in the 20th century to develop eight scenarios (Fig. 13) for California coastal climate at finer spatial scales than those produced by GCMs to evaluate if projections of future mean regional climate have observed analogs in the historic climatic record. We found evidence that supports the use of our warmer (normal precipitation) scenario as equivalent to the mean annual temperature and annual total precipitation projected changes for California by the CMIP5 projections. This finding has significant implications because it suggests that we could potentially use this high resolution "equivalent" as an alternative to downscaled GCMs. Specifically, if this "equivalent" is used in SDMs, we can produce more reasonable and physically accurate estimates of the anticipated range shifts in response to projected climate change not only for redwoods but for all terrestrial species that live in the area of influence of the upwelling zone in California. Consistent with Hansen (2012), CMIP5-RCP4.5 data confirmed that our anomalous years for the warmer (normal precipitation) scenario are in the vicinity of the projections for the 2020's and 2030's, which limits projections based on historic climate variability to the next few decades. We could also

note that for changes in that time frame, our equivalent scenario predicted a suitable habitat shift of 201 km to the north-northwest, and a contraction of 50% of that predicted by the "normal" conditions.

Key issues relevant to conservation planning and resource management that can be addressed through examination of the ensemble SDMs in response to anomalous conditions using our method include: identification of climatically highly stable and highly variable zones and identification of new conditions for zones that are predicted to change. In conservation planning, addressing these issues can inform the prioritization of areas that have higher degrees of natural resilience (i.e., refugia) to climate change. Also, areas that are outside the current range but that are predicted to become climatically suitable could be used for translocation efforts under our projected scenario of climate change. In resource management, the prospect of using our methods to evaluate the response of economically important crops to changing climatic conditions in the coast of California is also of great interest (Diffenbaugh et al., 2011). One example is the highly profitable variety of *Vitis vinifera,* known as Pinot noir, that shares the same bioclimate of redwoods, and for which there has been mounting pressure from the wine industry to convert current redwood lands to vineyards (Dimson, 2012). The areas that, in our models, are identified as climatically suitable today but that do not include redwoods, can be used to find a win-win solution among the wine industry and natural resource managers by providing information on where current Pinot noir grape production could be expanded without affecting redwoods. The areas identified as climatically suitable under the scenario that parallels IPCC projected change can also be use as a planning tool for new vineyard establishment in anticipation of future change. Moreover, crops are not limited by the dispersal abilities of species in natural ecosystems and can take full advantage of distant and fragmented suitable habitat as it becomes available in the future. Still, this type of analysis will require more detailed time series analysis (i.e., month to month variability) that can provide information about seasonality in agriculture.

Conservation planning frameworks are more often required to incorporate a better understanding of the spatio-temporal relationships between actual biodiversity and climate. New and interesting avenues of research can be developed if dendrochronologies are fully integrated with SDMs based on time series climatic data. Tree-rings are the log-book of the long-term climatic conditions that occur across species range providing the raw measures of the species physiological responses to the environment at a superior spatial density than any current or future weather station network. Integrated with hypothesis of species distributions, the dendrochronology climate records can help us understand fundamental questions about the spatial variability of redwoods' sensitivity to natural climatic variability and response to persistent extreme climatic conditions. Future research should also focus on dynamically updating our projections as the future unfolds, feeding our models with improved and novel species occurrences datasets (e.g., citizen science observations) and more robust time series data that can provide the fuel for successful adaptive management strategies.

We produced a multi-scenario assessment of potential climate change impacts to redwood forest that can inform resource managers about conservation opportunities at 800 m spatial resolution. While natural resource managers are continually seeking greater spatial precision in the projections (McPherson et al., 2006), there is still a lag in how they can best incorporate the uncertainty associated with these fine spatial resolution projections. Species distributions model outputs can vary substantially and while our research addresses the issue of how climate manifests at local scales, our method doesn't eliminate uncertainties associated with fine-scale interpolated climate data (Fernandez et al., 2013), which are also models with limitations (Franklin 2009). As a consequence, the resulting mapped representations of projected suitable habitat should be carefully interpreted. Ways to estimate the uncertainty due to interpolation include weighting the pixels by how far they are from the closest weather stations, or by how heterogeneous the landscape is using a finer resolution digital elevation model (Fernandez et al., 2013). Also very important but rarely considered in ecological forecasts is how the projected climatically suitable habitats will interact with non-climatic factors such as land-use change. Habitat degradation and land conversion play critical roles in estimating suitable habitat stability and should be incorporated in the projections.

CONCLUSION

With estimates of the number of living species in the planet ranging between 2 to 8 million (Costello et al., 2013), we acknowledge that we have little idea about the actual number and that we know even less about their actual spatial distributions. However, we do know from paleo-records (e.g., Blois and Hadly, 2007; Gonzáles-Carranza et al., 2012) that climate has a fundamental influence on where species are distributed in space, and also that differences between climates do not need to be so dramatic for us to see variation in species spatial organization.

The evidence of human induced climate change is now unequivocal. Between 20 to 30 percent of species will be at increased risk of extinction if global warming exceeds 1.9^{o}C, and that the number could increase to 40 to 70 percent if global temperature exceed 6.3^{o}C (IPCC, 2007). With a projected increase in temperatures of ~2^{o}C by midcentury, regardless of the postulated emission scenario, it is critical that we try to understand the impacts of climate change on biodiversity. However, the mechanisms by which climate change can impact biological populations are multiple and remain only partially unraveled, making it extremely difficult to predict what impacts on biodiversity will be in the future.

Species distribution modeling is a useful tool for describing the climate in which a species can live. Using this technique, it is possible to draw on a map locations predicted to have favorable conditions for the species under a scenario of future climate change. Therefore, a species distribution model simply creates a hypothesis of where the climate conditions that a species inhabit at the current time will be in the future, given assumptions about greenhouse gas emissions and climate model abilities to represent local climate change. However, a species may not experience the full range of climatic conditions that it can tolerate. SDMs work under the assumption that the species-climate relationship will remain stable trough time. Also, SDMs assume that all the individuals of a species or population are identical and will respond in the same manner to changing conditions. Predictions from SDMs are, therefore, far from perfect and difficult to interpret. Yet the approach provides us with a first estimate of the extinction risk problem. Only equipped with a clear understanding of how the models work we can move forward to look at the result with a critical eye.

The argument that connects the three chapters in my dissertation is related to the spatio-temporal mismatch between species observations and environmental data. This issue is relevant to species distribution modeling, and can be traced back to the two main data models created to store, display and analyze digital information: vector and raster data models. Both models have advantages and limitations that have been extensively reviewed elsewhere (e.g., Zhang and Goodchild 2002). The assumptions behind each of these models are different, and until very recently it was still difficult to work with both types of data models inside the same

GIS environment without applying a data transformation. The fundamental difference between these two data models relevant to our discussion, is that while the vector model can scale without any degradation, whereas, the raster model quality is completely dependent on the resolution (grid cell size). These data models are employed inside geographic information systems software to store, analyze and display different types of spatial information that can range from discrete data, which can relatively accurately be represented with points and lines that connect these points (i.e., vector data model), to continuous data that require some degree of discretization before it can be incorporated into a GIS (i.e., raster data model). Relevant to our problem is that the two sources of input data in SDM (i.e., species observations and environmental variables) are collected in their raw form using the vector data model with attributes that are measured from *x*, *y*, *z* coordinates, where the *x* and *y* represent the geographic position and *z* represents time.

Environmental information (*e*) is continuous data by nature (e.g., temperature and precipitation). However, due to practical limitations, it is measured from a relatively small sample of observations in the form of points (i.e., discrete). Also important is that for each *x*, *y* coordinate that represents a point, there are in general multiple *e* sample values collected at different *z* times. Despite their relative low spatial density, they provide us with a general view of the environmental conditions at different locations. Based on the idea that near things are more related than distant things (Tobler, 1970), we can use the *x*, *y* coordinates for the places from where we know the value of *e,* and interpolate the conditions for places lacking measurements. The result is a gridded layer where each grid cell has a value that represents the estimated value of *e* based on the closest samples. Applying this process to environmental information is common but it has consequences that affect its quality and accuracy. One consequence is that the data is transformed from the vector model into a raster model; a transformation process that requires the selection of a grid cell size. A second consequence is that in general the value that is interpolated is an average of multiple *e* values over *z* and its representativeness is determined by the longevity and consistency of the measure over time.

In the case of species occurrences (*s*), the spatial density of *x*, *y* measurements is several orders of magnitude lower than the density of environmental information. It also is rare to have more than one measure of *s* for every *x*, *y* sample data point over time (but see: Smith et al., 2013). Moreover, the *x*, *y* coordinates for *s* are rarely the same *x*, *y* coordinates for *e* samples. Current SDMs therefore correlate *e* (in the transformed raster model) with *s* in vector model *x*, *y* coordinates. The inferences that SDMs make about species climatic preferences carry some degree of uncertainty because *e* and *s* are not measured at the same *x*, *y* and *z*. Integrating species observations and climatic variables that are not measured at the same spatial and temporal resolution thus hinders our ability to forecast species range shifts and expansions in response to global.

Where sufficient occurrence data in the invaded range is available, the inclusion of an additional scale of climatic variability can enhance ecological understanding of the invasion events. However, when invaded range occurrence data is not available, the most

conservative approach would be to use only monthly climatic variability (Fernández et al., 2012). These findings have also very important implications for the SDM community in general, especially for those of us interested in projecting species distributions in time. From Global Climate Model simulations, we know that future conditions will not only shift in their mean values, but that they will also increase their variability from one year to the next. This means that if two similar locations under future conditions show no significant change in average, they might still present differences in their variability and, thus, species can perceive these two locations as very different environments.

Understanding species' response to multiple scales of climate variability in the past can help us understand how they will respond to changes in different scales of climate variability in the future. Not only relevant to species distribution modeling but in a more broad ecological sense, a measure that captures the relative variability from one year to the next in a map can help us identify populations within a species that live under a relative higher degree of climatic stress, and perhaps use this information as a surrogate for their adaptive capacity to future changing conditions. Also, derived higher level statistics, such as analysis of extremes, can provide some information that can be used as proxy for species' physiological limitations.

The second chapter of my thesis confirms the importance of characterizing uncertainty in interpolated climate layers derived from weather stations. My results identify three clear sources of uncertainty that have an effect on model performance: environmental heterogeneity, inter-annual climatic variability and distance to the closet weather station (Fernández et al., 2013). Nonetheless, I acknowledge that these are not all the possible sources of uncertainty in interpolated products. Two elements related to the weather stations that should also be considered in future studies are the longevity (i.e., for how long the station has been collecting records) and the consistency (i.e., gaps in the data and also changes in instrumentation that can affect the measurements). Also important, is that I analyzed the three sources of uncertainty as independent elements, and there are circumstances where one source of uncertainty can enhance another one. For example, areas of high topographic heterogeneity, in some regions, are also remote areas with very low accessibility and, as a consequence, very low density of weather stations. Future research should evaluate the interactions between these variables.

Environmental heterogeneity and inter-annual climatic variability can be derived in multiple ways and without necessarily having access to the original weather stations data that went into the creation of the final interpolated climatic products. However, for the distance to the closest weather station variable, which happens to be the one that has the highest effect on model performance, the information about weather stations locations used in interpolated products is crucial to identify regions where SDM users need to be particularly careful when making a decision on a model. A particular problem in developing countries is that most of the weather station information is not in digital form. If it is a challenge to obtain the locations of weather stations in these countries, is even more challenging to incorporate the records of those stations in interpolated products. The SDM community and the conservation community that often has more presence in

these countries should make an effort to promote the digitalization and sharing of these records.

The results of the third chapter of my thesis suggest that we could potentially use this high resolution "equivalent" as an alternative to downscaled GCMs in species distribution modeling and produce more reasonable and physically accurate estimates of the anticipated range shifts in response to projected climate change. When I compare the scenarios I developed to data from the CMIP5 the results showed that the warmer (normal precipitation) scenario is in the vicinity of the projections for the 2020's and 2030's, which limits the use of our method to this time frame. Regions similar to California, where local manifestations of global climate depend on ocean-atmosphere dynamics are not rare around the world and our method provides an opportunity to understand the links between upwelling pulses and the terrestrial ecosystems that surround them, as a step towards its conservation and appropriate management.

Important is that not all future climatic downscaled datasets are compatible with every question, taxa or region and sometimes spatial resolution is not the best metric to measure the appropriateness of a gridded datasets to be applied to answer certain question. The SDM community needs to engage climate modelers that can guide the selection of downscaled gridded datasets used to project species distribution in the future.

Despite my findings, several methodological issues remain to be explored and solved in order to provide the methodological rigor that the species distribution modeling field highly needs. The preceding chapters that constitute this thesis abound with challenges for future work and new directions, hopefully as a consequence of a promising area of research. Some of the major future challenges that I perceive include: (1) the need to improve our species occurrence records, (2) the need to improve global environmental datasets, and finally (3) how to assess quality in aggregated datasets.

The first point is related to the fact that for most of the species today we are forced to build 100% complete maps by using a sample of less than 0.1%. Although field observations are usually very expensive to obtain, it is very important to promote the maintenance and constant update of biodiversity repositories. Despite the initiatives to integrate the efforts of institutions that collect and host primary biodiversity data around the world (GBIF), we are still far away from having a global network of biodiversity that can not only serve as a data archive but also synthesize spatial and taxonomic information to provide yearly assessments of potential sites for new expeditions and revisits that are based on taxonomic, spatial and temporal coverage.

The second point is related to the unique challenges that modeling the effect of climate change and invasive species present. For the U.S., the only sub-kilometer climatic time series dataset that exists at a continental scale is PRISM, which becomes useless if the range of your species is shifting into Canada, even worse if the species of interest is an invasive from a different country. Climate change range shifts and invasive species range expansions are the two applications of SDM that affect biodiversity regardless of political

boundaries or data boundary limitations. Solving global problems require global datasets created with the same quality standards. However, this will require a truly global effort to promote sharing and availability of global weather station data. Future research should focus on the creation of high resolution time series climatic data for the entire planet.

Related to the previous two is the issue of aggregated datasets. It is very unusual in species distribution modeling to use data that was collected specifically with the objective to model the distribution of the species. Most of the time, modelers harvest and query multiple online databases and resources that they are familiar with at the time that is allocated for this task inside a project. However, they differ in quality from one dataset the next in projects that sometimes take advantage of hundreds of datasets (Fernandez et. al., 2012) is not transparent. Although there are some efforts to standardize the characterization of spatial uncertainty in museum collections the true is that once those georeferenced museum collections are integrated into a GIS the uncertainty associated with the latitude and longitude is often forgotten.

Clearly much more work is needed and although the preceding chapters in my dissertation are restricted to a few aspects of the field of species distribution modeling I hope I have achieved nonetheless a step toward a better understanding of the spatial distribution of species in the planet. I believe that spatio-temporal analysis of biodiversity aided with technological advances and a solid conceptual development has the potential to bridge disciplines and to serve as a framework for truly cross disciplinary research and synthesis.

REFERENCES

Abatzoglou, J. T. and T. J. Brown. 2012. A comparison of statistical downscaling methods suited for wildfire applications. International Journal of Climatology **32**:772-780.

Abatzoglou, J. T., K. T. Redmond, and L. M. Edwards. 2009. Classification of regional climate variability in the state of California. Journal of Applied Meteorology and Climatology **48**:1527-1541.

Aguilar, F. J., M. A. Aguilar, F. Agüera, and J. Sánchez. 2006. The accuracy of grid digital elevation models linearly constructed from scattered sample data. International Journal of Geographical Information Science **20**:169-192.

Allouche, O., A. Tsoar, and R. Kadmon. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology **43**:1223-1232.

Angert, A. L. 2009. The niche, limits to species' distributions, and spatiotemporal variation in demography across the elevation ranges of two monkeyflowers. Proceedings of the National Academy of Sciences of the United States of America **106**:19693-19698.

Araujo, M. and C. Rahbek. 2006. How does climate change affect biodiversity? Science **313**:1396-1397.

Araujo, M. B. and A. Guisan. 2006. Five (or so) challenges for species distribution modelling. Journal of Biogeography **33**:1677-1688.

Araujo, M. B. and R. G. Pearson. 2005. Equilibrium of species' distributions with climate. Ecography **28**:693-695.

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

Bakun, A. 1990. Global climate change and intensification of coastal ocean upwelling. Science **247**:198-201.

Bakun, A., D. B. Field, A. Redondo-Rodriguez, and S. J. Weeks. 2010. Greenhouse gas, upwelling-favorable winds, and the future of coastal ocean upwelling ecosystems. Global Change Biology **16**:1213-1228.

Barron, J. A., L. Heusser, T. Herbert, and M. Lyle. 2003. High-resolution climatic evolution of coastal northern California during the past 16,000 years. Paleoceanography **18**:1020.

Barve, N., V. Barve, A. Jimenez-Valverde, A. Lira-Noriega, S. P. Maher, A. T. Peterson, J. Soberon, and F. Villalobos. 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecological Modelling **222**:1810-1819.

Baveye, P. 2009. Uncertainties in environmental modelling and consequences for policy making. Springer, Dordrecht, The Netherlands.

Beale, C. M. and J. J. Lennon. 2012. Incorporating uncertainty in predictive species distribution modelling. Philosophical Transactions of the Royal Society B: Biological Sciences **367**:247-258.

Beaumont, L. J., R. V. Gallagher, W. Thuiller, P. O. Downey, M. R. Leishman, and L. Hughes. 2009. Different climatic envelopes among invasive populations may lead to underestimations of current and future biological invasions. Diversity and Distributions **15**:409-420.

Beyer, H. 2010. Geospatial modelling environment. Spatial Ecology LLC, Toronto, Canada.

BirdLife. 2012. BirdLife International Species factsheets. BirdLife International, Cambridge, UK.

Blois, J. and E. Hadly. 2007. Mammalian response to quaternary climate change in Northern California. Journal of Vertebrate Paleontology **27**:49A-50A.

Broennimann, O. and A. Guisan. 2008. Predicting current and future biological invasions: both native and invaded ranges matter. Biology Letters **4**:585-589.

Broennimann, O., U. A. Treier, H. Müller-Schärer, W. Thuiller, A. T. Peterson, and A. Guisan. 2007. Evidence of climatic niche shift during biological invasion. Ecology Letters **10**:701-709.

Burgess, S. and T. Dawson. 2004. The contribution of fog to the water relations of *Sequoia sempervirens* (D. Don): foliar uptake and prevention of dehydration. Plant, Cell & Environment **27**:1023-1034.

Burgman, M. A. and J. C. Fox. 2003. Bias in species range estimates from minimum convex polygons: implications for conservation and options for improved planning. Animal Conservation **6**:19-28.

Cadena, D. C. and B. A. Loiselle. 2007. Limits to elevational distributions in two species of emberizine finches: disentangling the role of interspecific competition, autoecology, and geographic variation in the environment. Ecography **30**:491-504.

Cadotte, M. W., S. M. McMahon, and T. Fukami. 2006. Conceptual ecology and invasion biology: reciprocal approaches to nature. Springer Dordrecht, The Netherlands.

Calder, W. A. 1994. When do hummingbirds use torpor in nature? Physiological Zoology **76**:1051-1076.

Carroll, C., J. R. Dunk, and A. Moilanen. 2010. Optimizing resiliency of reserve networks to climate change: multispecies conservation planning in the Pacific Northwest, USA. Global Change Biology **16**:891-904.

Castro, C. L., R. A. Pielke, and G. Leoncini. 2005. Dynamical downscaling: Assessment of value retained and added using the Regional Atmospheric Modeling System (RAMS). Journal of Geophysical Research: Atmospheres (1984-2012) **110**.

Chapman, A. and J. Wieczorek. 2006. Guide to Best Practices for Georeferencing. Global Biodiversity Information Facility, Copenhagen, Denmark.

Conover, W. J. 1999. Practical nonparametric statistics. John Wiley & Sons, New York, New York.

Cordero, E. C., W. Kessomkiat, J. Abatzoglou, and S. A. Mauget. 2011. The identification of distinct patterns in California temperature trends. Climatic change **108**:357-382.

Costello, M. J., R. M. May, and N. E. Stork. 2013. Can we name Earth's species before they go extinct? Science **339**:413-416.

Cox, G. W. 2004. Alien species and evolution: the evolutionary ecology of exotic plants,

animals, microbes, and interacting native species. Island Press, Washington D.C.
da Mata, R. A., R. Tidon, L. G. Cortes, P. De Marco Jr, and J. A. F. Diniz-Filho. 2010. Invasive and flexible: niche shift in the drosophilid *Zaprionus indianus* (Insecta, Diptera). Biological Invasions **12**:1231-1241.
Daly, C. 2006. Guidelines for assessing the suitability of spatial climate data sets. International Journal of Climatology **26**:707-721.
Daly, C., G. Taylor, W. Gibson, T. Parzybok, G. Johnson, and P. Pasteris. 2000. High-quality spatial climate data sets for the United States and beyond. Transactions of the ASAE-American Society of Agricultural Engineers **43**:1957-1962.
Daszak, P., C. Zambrana-Torrelio, T. Bogich, M. Fernandez, J. Epstein, K. Murray, and H. Hamilton. 2012. Interdisciplinary approaches to understanding disease emergence: The past, present, and future drivers of Nipah virus emergence. Proceedings of the National Academy of Sciences of the United States of America **110**:3681-3688.
de Oliveira, G., M. B. Araujo, T. F. Rangel, D. Alagador, and J. A. F. Diniz-Filho. 2012. Conserving the Brazilian semiarid (Caatinga) biome under climate change. Biodiversity and Conservation **21**:2913-2926.
Diaz, H. F. and V. Markgraf. 2000. El Niño and the Southern Oscillation: multiscale variability and global and regional impacts. Cambridge University Press, Cambridge, UK.
Diffenbaugh, N. S., M. A. White, G. V. Jones, and M. Ashfaq. 2011. Climate adaptation wedges: a case study of premium wine in the western United States. Environmental Research Letters **6**:024024.
Dimson, M. 2012. At the Corner of Redwood and Vine: Perspectives on Forest-to-Vineyard Conversion Conflicts in Sonoma County. Sonoma State University, Sonoma, California.
Dixon, P. 1993. The bootstrap and the jackknife: describing the precision of ecological indices. Pages 290-318 *in* S. M. Scheiner and J. Gurevitch, editors. Design and analysis of ecological experiments. Oxford University Press, New York, U.S.
Drew, C. A. 2011. Predictive species and habitat modeling in landscape ecology. Springer Verlag, London, UK.
Drut, M. S., J. A. Crawford, and M. A. Gregg. 1994. Brood habitat use by sage grouse in Oregon. Western North American Naturalist **54**:170-176.
Efron, B. 1982. The Jacknife, the Bootstrap and Other Resampling Plans. Society for Industrial and Applied Mathematics, Philadelphia, U.S.
Elith, J. and C. H. Graham. 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., H. Graham, P. Anderson, M. Dudik, S. Ferrier, A. Guisan, J. Hijmans, F. Huettmann, R. Leathwick, and A. Lehmann. 2006. Novel methods improve prediction of species distributions from occurrence data. Ecography **29**:129-151.
Elith, J., M. Kearney, and S. Phillips. 2010. The art of modelling range-shifting species. Methods in Ecology and Evolution **1**:330-342.
Elith, J. and J. Leathwick. 2009. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology Evolution and

Systematics **40**:677-697.
Elith, J. and J. R. Leathwick. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. Annual Review of Ecology Evolution and Systematics **40**:677-697.
Elith, J., S. Phillips, T. Hastie, M. Dudík, Y. E. Chee, and C. Yates. 2011. A statistical explanation of MaxEnt for ecologists. Ecology **45**:1372-1381.
Ellis, A. W., G. B. Goodrich, and G. M. Garfin. 2010. A hydroclimatic index for examining patterns of drought in the Colorado River Basin. International Journal of Climatology **30**:236-255.
Ellis, A. W., T. W. Hawkins, R. Balling, and P. Gober. 2008. Estimating future runoff levels for a semi-arid fluvial system in central Arizona, USA. Climate Research **35**:227.
Everitt, B. 2005. An R and Splus Companion to Multivariate Analysis. Springer-Verlag, London, UK.
FAOCLIM-2. 2005. World-wide Agroclimatic Database. Food and Agriculture Organization. FAO, Rome, Italy.
Farr, T. G., P. A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, and L. Roth. 2007. The shuttle radar topography mission. Reviews of Geophysics **45**:RG2004.
Feeley, K. J. and M. R. Silman. 2011. Keep collecting: accurate species distribution modelling requires more collections than previously thought. Diversity and Distributions **17**:1132-1140.
Fernández, M., S. Blum, S. Reichle, Q. Guo, B. Holzman, and H. Hamilton. 2009. Locality uncertainty and the differential performance of four common niche-based modeling techniques. Biodiversity Informatics **6**:36-52.
Fernández, M., H. Hamilton, O. Alvarez, and Q. Guo. 2012. Does adding multi-scale climatic variability improve our capacity to explain niche transferability in invasive species? Ecological Modelling **246**:60-67.
Fernández, M., H. Hamilton, and L. M. Kueppers. 2013. Characterizing uncertainty in species distribution models derived from interpolated weather station data. Ecosphere **4**:5, art61.
Fielding, A. H. and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation **24**:38-49.
Fitzgerald, K. and D. M. Gordon. 2012. Effects of vegetation cover, presence of a native ant species, and human disturbance on colonization by Argentine ants. Conservation Biology **26**:525-538.
Flint, L. E. and A. L. Flint. 2012. Downscaling future climate scenarios to fine scales for hydrologic and ecological modeling and analysis. Ecological Processes **1**:1-15.
Fordham, D. A., H. R. Akcakaya, M. B. Araujo, J. Elith, D. A. Keith, R. Pearson, T. D. Auld, C. Mellin, J. W. Morgan, T. J. Regan, M. Tozer, M. J. Watts, M. White, B. A. Wintle, C. Yates, and B. W. Brook. 2012. Plant extinction risk under climate change: are forecast range shifts alone a good indicator of species vulnerability to global warming? Global Change Biology **18**:1357-1371.
Forsythe, W. C., E. J. Rykiel, R. S. Stahl, H. Wu, and R. M. Schoolfield. 1995. A model

comparison for daylength as a function of latitude and day of year. Ecological Modelling **80**:87-95.

Franklin, J., F. W. Davis, M. Ikegami, A. D. Syphard, L. E. Flint, A. L. Flint, and L. Hannah. 2013. Modeling plant species distributions under future climates: how fine scale do climate projections need to be? Global Change Biology **19**:473-483.

Franklin, J. and J. Miller. 2009. Mapping species distributions: spatial inference and prediction. Cambridge University Press, Cambridge, UK.

Freas, K. and P. R. Kemp. 1983. Some relationships between environmental reliability and seed dormancy in desert annual plants. The Journal of Ecology **71**:211-217.

Fuller, T., H. A. Thomassen, P. M. Mulembakani, S. C. Johnston, J. O. Lloyd-Smith, N. K. Kisalu, T. K. Lutete, S. Blumberg, J. N. Fair, and N. D. Wolfe. 2011. Using remote sensing to map the risk of human monkeypox virus in the Congo Basin. EcoHealth **8**:14-25.

Ghil, M. 2002. Natural Climatic Variability. Pages 544–549 *in* T. Munn, editor. Encyclopedia of global environmental change. John Wiley & Sons, Chichester, UK.

Giesecke, T., P. A. Miller, M. T. Sykes, A. E. Ojala, H. Seppä, and R. H. Bradshaw. 2010. The effect of past changes in inter-annual temperature variability on tree distribution limits. Journal of Biogeography **37**:1394-1405.

Giuliano, W. and R. Lutz. 1993. Quail and rain: what's the relationship. Pages 64-68 *in* Proceedings of the National Quail Symposium.

Gonzalez, C., O. Wang, S. E. Strutz, C. Gonzalez-Salazar, V. Sanchez-Cordero, and S. Sarkar. 2010. Climate change and risk of leishmaniasis in North America: predictions from ecological niche models of vector and reservoir species. PLoS neglected tropical diseases **4**:e585.

González-Carranza, Z., H. Hooghiemstra, and M. I. Vélez. 2012. Major altitudinal shifts in Andean vegetation on the Amazonian flank show temporary loss of biota in the Holocene. The Holocene **22**:1227-1241.

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

Graham, C. H., S. R. Ron, J. C. Santos, C. J. Schneider, and C. Moritz. 2004. Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. Evolution **58**:1781-1793.

Guo, Q., W. Li, H. Yu, and O. Alvarez. 2010. Effects of topographic variability and lidar sampling density on several DEM interpolation methods. Photogrammetric Engineering and Remote Sensing **76**:701-712.

Hamon, W. R. 1963. Computation of direct runoff amounts from storm rainfall. International Association of Hydrological Sciences.

Hannah, L., P. R. Roehrdanz, M. Ikegami, A. V. Shepard, M. R. Shaw, G. Tabor, L. Zhi, P. A. Marquet, and R. J. Hijmans. 2013. Climate change, wine, and conservation. Proceedings of the National Academy of Sciences **110**:6907-6912.

Hansen, J., M. Sato, and R. Ruedy. 2012. Perception of climate change. Proceedings of the National Academy of Sciences of the United States of America **109**:E2415-E2423.

Hegel, T. M., S. A. Cushman, J. Evans, and F. Huettmann. 2010. Current state of the art for statistical modelling of species distributions. Pages 273-311 *in* S. Cushman and F. Huettmann, editors. Spatial Complexity, Informatics, and Wildlife Conservation. Springer, Tokyo, Japan.

Heikkinen, R. K., M. Luoto, M. B. Araujo, R. Virkkala, W. Thuiller, and M. T. Sykes. 2006. Methods and uncertainties in bioclimatic envelope modelling under climate change. Progress in Physical Geography **30**:751-777.

Helmuth, B., C. D. Harley, P. M. Halpin, M. O'Donnell, G. E. Hofmann, and C. A. Blanchette. 2002. Climate change and latitudinal patterns of intertidal thermal stress. Science **298**:1015-1017.

Hidalgo, H. G., M. D. Dettinger, and D. R. Cayan. 2008. Downscaling with constructed analogues: Daily precipitation and temperature fields over the United States. California Climate Change Center:48.

Hijmans, R., S. Cameron, J. Parra, P. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology **25**:1965-1978.

Holt, A. C., D. J. Salkeld, C. L. Fritz, J. R. Tucker, and P. Gong. 2009. Spatial analysis of plague in California: niche modeling predictions of the current distribution and potential response to climate change. International Journal of Health Geographics **8**:38.

Hortal, J., A. Jiménez Valverde, J. F. Gómez, J. M. Lobo, and A. Baselga. 2008. Historical bias in biodiversity inventories affects the observed environmental niche of the species. Oikos **117**:847-858.

Howell, S. N. G. and S. Webb. 1995. A guide to the birds of Mexico and northern Central America. Oxford University Press, New York, New York.

Huston, M. A. 2002. Introductory essay: critical issues for improving predictions. Pages 7-21 *in* J. M. Scott, P. Heglund, M. Morrison, J. Haufler, M. Raphael, W. Wall, and F. Samson, editors. Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Covelo, California.

Hutchins, D., C. Hare, R. Weaver, Y. Zhang, G. Firme, G. DiTullio, M. Alm, S. Riseman, J. Maucher, and M. Geesey. 2002. Phytoplankton iron limitation in the Humboldt Current and Peru Upwelling. Limnology and Oceanography:997-1011.

Hutchinson, G. E. 1957. Concluding remarks. Pages 415-427 *in* Cold Spring Harbor Symposia on Quantitative Biology. Cold Spring Harbor Laboratory Press.

IUCN. 2012. The IUCN Red List of Threatened Species Version 2012.1. International Union for Conservation of Nature and Natural Resources, Cambridge, UK.

Jackson, S. T., J. L. Betancourt, R. K. Booth, and S. T. Gray. 2009. Ecology and the ratchet of events: climate variability, niche dimensions, and species distributions. Proceedings of the National Academy of Sciences of the United States of America **106**:19685-19692.

Jeschke, J. M. and D. L. Strayer. 2008. Usefulness of bioclimatic models for studying climate change and invasive species. Annals of the New York Academy of Sciences **1134**:1-24.

Jiguet, F. d. r., L. Brotons, and V. Devictor. 2011. Community responses to extreme climatic conditions. Current Zoology **57**:406-413.

Johnsgard, P. A. 1983. The hummingbirds of North America. Smithsonian Institution Press, Washington D.C.

Johnsgard, P. A. and H. Jones. 1988. The quails, partridges, and francolins of the world. Oxford University Press, New York, New York.

Johnson, G. L., C. Daly, G. H. Taylor, and C. L. Hanson. 2000. Spatial variability and interpolation of stochastic weather simulation model parameters. Journal of Applied Meteorology **39**:778-796.

Kamilar, J. M. 2009. Environmental and geographic correlates of the taxonomic structure of primate communities. American Journal of Physical Anthropology **139**:382-393.

Kamino, L. H. Y., J. R. Stehmann, S. Amaral, P. De Marco, T. F. Rangel, M. F. de Siqueira, R. De Giovanni, and J. Hortal. 2012. Challenges and perspectives for species distribution modelling in the neotropics. Biology Letters **8**:324-326.

Klausmeyer, K. R. and M. R. Shaw. 2009. Climate change, habitat loss, protected areas and the climate adaptation potential of species in Mediterranean ecosystems worldwide. PLoS ONE **4**:e6392.

Knutti, R. and J. Sedlacek. 2013. Robustness and uncertainties in the new CMIP5 climate model projections. Nature Climate Change **3**:369-373.

Kreft, H. and W. Jetz. 2007. Global patterns and determinants of vascular plant diversity. Proceedings of the National Academy of Sciences of the United States of America **104**:5925-5930.

Kremen, C., A. Cameron, A. Moilanen, S. Phillips, C. Thomas, H. Beentje, J. Dransfield, B. Fisher, F. Glaw, and T. Good. 2008. Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools. Science **320**:222-226.

Kriticos, D. J. and A. Leriche. 2010. The effects of climate data precision on fitting and projecting species niche models. Ecography **33**:115-127.

Kueppers, L. M., M. A. Snyder, L. C. Sloan, E. S. Zavaleta, and B. Fulfrost. 2005. Modeled regional climate change and California endemic oak ranges. Proceedings of the National Academy of Sciences of the United States of America **102**:16281-16286.

LaDochy, S., R. Medina, and W. Patzert. 2007. Recent California climate variability: spatial and temporal patterns in temperature trends. Climate Research **33**:159-169.

Lebassi, B., J. Gonzalez, D. Fabris, E. Maurer, N. Miller, C. Milesi, P. Switzer, and R. Bornstein. 2009. Observed 1970-2005 Cooling of Summer Daytime Temperatures in Coastal California. Journal of Climate **22**:3558-3573.

Lenoir, J., J.-C. Gégout, A. Guisan, P. Vittoz, T. Wohlgemuth, N. E. Zimmermann, S. Dullinger, H. Pauli, W. Willner, and J.-C. Svenning. 2010. Going against the flow: potential mechanisms for unexpected downslope range shifts in a warming climate. Ecography **33**:295-303.

Levsen, N. D., P. Tiffin, and M. S. Olson. 2012. Pleistocene speciation in the genus *Populus* (Salicaceae). Systematic biology **61**:401-412.

Li, W., Q. Guo, and C. Elkan. 2011. Can we model the probability of presence of species without absence data? Ecography **34**:1096-1105.

Loarie, S. R., B. E. Carter, K. Hayhoe, S. McMahon, R. Moe, C. A. Knight, and D. D. Ackerly. 2008. Climate Change and the Future of California's Endemic Flora.

PLoS ONE **3**.
Lobo, J. M., A. Jiménez Valverde, and J. Hortal. 2010. The uncertain nature of absences and their importance in species distribution modelling. Ecography **33**:103-114.
Lobo, J. M., A. Jiménez-Valverde, and R. Real. 2008. AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography **17**:145-151.
Lockwood, J., M. Hoopes, and M. Marchetti. 2009. Invasion ecology. Wiley-Blackwell, Oxford, UK.
Loiselle, B. A., P. M. Jørgensen, T. Consiglio, I. Jiménez, J. G. Blake, L. G. Lohmann, and O. M. Montiel. 2008. Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? Journal of Biogeography **35**:105-116.
Lomolino, M. V. 2010. Biogeography. 4th edition. Sinauer Associates, Sunderland, Massachusetts.
Lorenz, E. N. 1969. Atmospheric predictability as revealed by naturally occurring analogues. Journal of the Atmospheric Sciences **26**:636-646.
Lowe, S., M. Browne, S. Boudjelas, and M. De Poorter. 2000. 100 of the world's worst invasive alien species: a selection from the global invasive species database. Invasive Species Specialist Group Auckland, New Zealand.
Mann, K. H. and J. R. N. Lazier. 1996. Dynamics of marine ecosystems. Third edition. Blackwell Science Cambridge, Malden, Massachusetts.
Marmion, M., M. Luoto, R. K. Heikkinen, and W. Thuiller. 2009. The performance of state-of-the-art modelling techniques depends on geographical distribution of species. Ecological Modelling **220**:3512-3520.
Martin, T. E. and J. J. Roper. 1988. Nest predation and nest-site selection of a western population of the Hermit Thrush. Condor **90**:51-57.
McInerny, G. J. and D. W. Purves. 2011. Fine-scale environmental variation in species distribution modelling: regression dilution, latent variables and neighbourly advice. Methods in Ecology and Evolution **2**:248-257.
Mitchell, T. D. and P. D. Jones. 2005. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. International Journal of Climatology **25**:693-712.
Mooney, H. A. and R. J. Hobbs. 2000. Invasive species in a changing world. Island Press, Washington, D.C.
Moritz, C. 2002. Strategies to protect biological diversity and the evolutionary processes that sustain it. Systematic biology **51**:238-254.
Naimi, B., A. K. Skidmore, T. A. Groen, and N. A. S. Hamm. 2011. Spatial autocorrelation in predictors reduces the impact of positional uncertainty in occurrence data on species distribution modelling. Journal of Biogeography **38**:1497-1509.
NatureServe. 2012. NatureServe Explorer: an online encyclopedia of life Version 7.1. NatureServe, Arlington, Virginia.
Nix, H. A. 1986. A biogeographic analysis of Australian elapid snakes. Atlas of elapid snakes of Australia **7**:4-15.
Noss, R. F. 2000. The redwood forest: history, ecology, and conservation of the coast

redwoods. Island Press, Washington, D.C.
O'Brien, T. A., L. C. Sloan, P. Y. Chuang, I. C. Faloona, and J. A. Johnstone. 2012. Multidecadal simulation of coastal fog with a regional climate model. Climate Dynamics:1-12.
Osmond, B., G. Ananyev, J. Berry, C. Langdon, Z. Kolber, G. H. Lin, R. Monson, C. Nichol, U. Rascher, U. Schurr, S. Smith, and D. Yakir. 2004. Changing the way we think about global change research: scaling up in experimental ecosystem science. Global Change Biology **10**:393-407.
Paltineanu, C., I. Mihailescu, Z. Prefac, C. Dragota, F. Vasenciuc, and N. Claudia. 2009. Combining the standardized precipitation index and climatic water deficit in characterizing droughts: a case study in Romania. Theoretical and Applied Climatology **97**:219-233.
Pearman, P. B., A. Guisan, O. Broennimann, and C. F. Randin. 2008. Niche dynamics in space and time. Trends in Ecology & Evolution **23**:149-158.
Pearson, R. 2011. Driven to extinction: the impact of climate change on biodiversity. Sterling New York.
Peters, G. P., R. M. Andrew, T. Boden, J. G. Canadell, P. Ciais, C. Le Quéré, G. Marland, M. R. Raupach, and C. Wilson. 2012. The challenge to keep global warming below 2 degrees C. Nature Climate Change **3**.
Peterson, A. T. and Y. Nakazawa. 2008. Environmental data sets matter in ecological niche modelling: an example with *Solenopsis invicta* and *Solenopsis richteri*. Global Ecology and Biogeography **17**:135-144.
Peterson, A. T., M. Papes, and M. Eaton. 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. Ecography **30**:550-560.
Peterson, A. T. and C. R. Robins. 2003. Using ecological-niche modeling to predict Barred Owl invasions with implications for Spotted Owl conservation. Conservation Biology **17**:1161-1165.
Peterson, A. T., J. Soberon, R. G. Pearson, R. P. Anderson, E. Martinez-Meyer, M. Nakamura, and M. B. Araujo. 2011. Ecological Niches and Geographic Distributions (MPB-49). Princeton University Press, Princeton, New Jersey.
Peterson, A. T. and D. A. Vieglais. 2001. Predicting species invasions using ecological niche modeling: New approaches from bioinformatics attack a pressing problem. Bioscience **51**:363-371.
Phillips, S. J. 2008. Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson et al.(2007). Ecography **31**:272-278.
Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling **190**:231-259.
Phillips, S. J. and M. Dudik. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography **31**:161–175.
Picard, R. R. and R. D. Cook. 1984. Cross-validation of regression models. Journal of the American Statistical Association **79**:575-583.
Purvis, A., S. A. Fritz, J. Rodríguez, P. H. Harvey, and R. Grenyer. 2011. The shape of mammalian phylogeny: patterns, processes and scales. Philosophical Transactions

of the Royal Society B: Biological Sciences **366**:2462-2477.
Pyke, C. R. 2005. Interactions between habitat loss and climate change: Implications for fairy shrimp in the Central Valley ecoregion of California, USA. Climatic change **68**:199-218.
Randin, C., T. Dirnbock, S. Dullinger, N. Zimmermann, M. Zappa, and A. Guisan. 2006. Are niche-based species distribution models transferable in space? Journal of Biogeography **33**:1689-1703.
Raxworthy, C. J., C. M. Ingram, N. Rabibisoa, and R. G. Pearson. 2007. Applications of ecological niche modeling for species delimitation: A review and empirical evaluation using day geckos (*Phelsuma*) from Madagascar. Systematic biology **56**:907-923.
Reside, A. E., J. J. VanDerWal, A. S. Kutt, and G. C. Perkins. 2010. Weather, not climate, defines distributions of vagile bird species. PLoS ONE **5**:e13569.
Robb, G. N., R. A. McDonald, D. E. Chamberlain, and S. Bearhop. 2008. Food for thought: supplementary feeding as a driver of ecological change in avian populations. Frontiers in Ecology and the Environment **6**:476-484.
Rödder, D. and S. Lötters. 2009. Niche shift versus niche conservatism? Climatic characteristics of the native and invasive ranges of the Mediterranean house gecko (*Hemidactylus turcicus*). Global Ecology and Biogeography **18**:674-687.
Rödder, D. and S. Lötters. 2010. Explanative power of variables used in species distribution modelling: an issue of general model transferability or niche shift in the invasive Greenhouse frog (*Eleutherodactylus planirostris*). Naturwissenschaften **97**:781-796.
Rodriguez, J. P., L. Brotons, J. Bustamante, and J. Seoane. 2007. The application of predictive modelling of species distribution to biodiversity conservation. Diversity and Distributions **13**:243-251.
Rosell, J. A., M. E. Olson, R. Aguirre-Hernandez, and F. J. Sanchez-Sesma. 2012. Ontogenetic modulation of branch size, shape, and biomechanics produces diversity across habitats in the *Bursera simaruba* clade of tropical trees. Evolution & Development **14**:437-449.
Rota, C. T., R. J. Fletcher, J. M. Evans, and R. L. Hutto. 2011. Does accounting for imperfect detection improve species distribution models? Ecography **34**:659-670.
Roura-Pascual, N., L. Brotons, A. T. Peterson, and W. Thuiller. 2009. Consensual predictions of potential distributional areas for invasive species: a case study of Argentine ants in the Iberian Peninsula. Biological Invasions **11**:1017-1031.
Roura-Pascual, N., C. Hui, T. Ikeda, G. Leday, D. M. Richardson, S. Carpintero, X. Espadaler, C. Gómez, B. Guénard, and S. Hartley. 2011. Relative roles of climatic suitability and anthropogenic influence in determining the pattern of spread in a global invader. Proceedings of the National Academy of Sciences of the United States of America **108**:220-225.
Roura-Pascual, N., A. V. Suarez, C. Gómez, P. Pons, Y. Touyama, A. L. Wild, and A. T. Peterson. 2004. Geographical potential of Argentine ants (*Linepithema humile* Mayr) in the face of global climate change. Proceedings of the Royal Society of London. Series B: Biological Sciences **271**:2527-2535.
Saab, V. A. and J. G. Dudley. 1998. Responses of cavity-nesting birds to stand-

replacement fire and salvage logging in ponderosa pine/Douglas-fir forests of southwestern Idaho. US Department of Agriculture, Forest Service, Ogden, UT.

Sauer, J. R., J. E. Fallon, and R. Johnson. 2003. Use of North American Breeding Bird Survey data to estimate population change for bird conservation regions. The Journal of wildlife management **67**:372-389.

Sauer, J. R., J. E. Hines, and J. Fallon. 2006. The North American breeding bird survey, results and analysis 1966-2007. USGS Patuxent Wildlife Research Center, Laurel, Madison.

Saupe, E., V. Barve, C. Myers, J. Soberon, N. Barve, C. Hensz, A. Peterson, H. Owens, and A. Lira-Noriega. 2012. Variation in niche and distribution model performance: The need for a priori assessment of key causal factors. Ecological Modelling **237**:11-22.

Sax, D. F., J. J. Stachowicz, J. H. Brown, J. F. Bruno, M. N. Dawson, S. D. Gaines, R. K. Grosberg, A. Hastings, R. D. Holt, and M. M. Mayfield. 2007. Ecological and evolutionary insights from species invasions. Trends in Ecology & Evolution **22**:465-471.

Schröder, B. 2008. Challenges of species distribution modeling belowground. Journal of Plant Nutrition and Soil Science **171**:325-337.

Schroeder, M. A. and C. E. Braun. 1993. Partial migration in a population of greater prairie-chickens in northeastern Colorado. The Auk **110**:21-28.

Simonin, K. A., L. S. Santiago, and T. E. Dawson. 2009. Fog interception by *Sequoia sempervirens* (D. Don) crowns decouples physiology from soil water deficit. Plant, Cell & Environment **32**:882-892.

Smith, A. B., M. J. Santos, M. S. Koo, K. Rowe, K. C. Rowe, J. L. Patton, J. D. Perrine, S. R. Beissinger, and C. Moritz. 2013. Evaluation of species distribution models by resampling of sites surveyed a century ago by Joseph Grinnell. Ecography **Online preview**.

Smolik, M., S. Dullinger, F. Essl, I. Kleinbauer, M. Leitner, J. Peterseil, L. Ä. Stadler, and G. Vogl. 2010. Integrating species distribution models and interacting particle systems to predict the spread of an invasive alien plant. Journal of Biogeography **37**:411-422.

Snyder, M., L. Sloan, N. Diffenbaugh, and J. Bell. 2003. Future climate change and upwelling in the California Current. Geophysical Research Letters **30**.

Soberón, J. M., J. B. Llorente, and L. Oñate. 2000. The use of specimen-label databases for conservation purposes: an example using Mexican Papilionid and Pierid butterflies. Biodiversity and Conservation **9**:1441-1466.

Steiner, F. M., B. C. Schlick-Steiner, J. VanDerWal, K. D. Reuther, E. Christian, C. Stauffer, A. V. Suarez, S. E. Williams, and R. H. Crozier. 2008. Combined modelling of distribution and niche in invasion biology: a case study of two invasive *Tetramorium* ant species. Diversity and Distributions **14**:538-545.

Stephenson, N. 1998. Actual evapotranspiration and deficit: biologically meaningful correlates of vegetation distribution across spatial scales. Journal of Biogeography **25**:855-870.

Stephenson, N. L. 1990. Climatic control of vegetation distribution: the role of the water balance. American Naturalist:649-670.

Stigall, A. L. 2012. Using ecological niche modelling to evaluate niche stability in deep time. Journal of Biogeography **39**:772–781.

Suarez-Seoane, S., P. E. Osborne, and A. Rosema. 2004. Can climate data from METEOSAT improve wildlife distribution models? Ecography **27**:629-636.

Synes, N. W. and P. E. Osborne. 2011. Choice of predictor variables as a source of uncertainty in continental-scale species distribution modelling under climate change. Global Ecology and Biogeography **20**:904-914.

Tabor, K. and J. W. Williams. 2010. Globally downscaled climate projections for assessing the conservation impacts of climate change. Ecological Applications **20**:554-565.

Terribile, L. and J. Diniz-Filho. 2010. How many studies are necessary to compare niche-based models for geographic distributions? Inductive reasoning may fail at the end. Brazilian Journal of Biology **70**:263-269.

Thomson, A. M., K. V. Calvin, S. J. Smith, G. P. Kyle, A. Volke, P. Patel, S. Delgado-Arias, B. Bond-Lamberty, M. A. Wise, and L. E. Clarke. 2011. RCP4. 5: a pathway for stabilization of radiative forcing by 2100. Climatic change **109**:77-94.

Thorn, J. S., V. Nijman, D. Smith, and K. A. I. Nekaris. 2009. Ecological niche modelling as a technique for assessing threats and setting conservation priorities for Asian slow lorises (Primates: *Nycticebus*). Diversity and Distributions **15**:289-298.

Thornthwaite, C. and J. Mather. 1955. The water balance. Centerton: Drexel Institute of Technology, Laboratory of Climatology. Publications in Climatology **3**.

Thuiller, W., S. b. Lavergne, C. Roquet, I. Boulangeat, B. Lafourcade, and M. B. Araujo. 2011. Consequences of climate change on the tree of life in Europe. Nature **470**:531-534.

Thuiller, W., D. M. Richardson, P. Pysek, G. F. Midgley, G. O. Hughes, and M. Rouget. 2005. Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. Global Change Biology **11**:2234-2250.

Tingley, M. W. and S. R. Beissinger. 2009. Detecting range shifts from historical species occurrences: new perspectives on old data. Trends in Ecology & Evolution **24**:625-633.

Titeux, N., D. Maes, M. Marmion, M. Luoto, and R. K. Heikkinen. 2009. Inclusion of soil data improves the performance of bioclimatic envelope models for insect species distributions in temperate Europe. Journal of Biogeography **36**:1459-1473.

Tobler, W. R. 1970. A computer movie simulating urban growth in the Detroit region. Economic Geography **46**:234-240.

Torregrosa, A., M. D. Taylor, L. E. Flint, and A. L. Flint. 2013. Present, future, and novel bioclimates of the San Francisco, California Region. PLoS ONE **8**:e58450.

Turner, W., S. Spector, N. Gardiner, M. Fladeland, E. Sterling, and M. Steininger. 2003. Remote sensing for biodiversity science and conservation. Trends in Ecology & Evolution **18**:306-314.

Vaclavik, T., J. A. Kupfer, and R. K. Meentemeyer. 2012. Accounting for multi-scale spatial autocorrelation improves performance of invasive species distribution

modelling (iSDM). Journal of Biogeography **39**:42-55.
VanDerWal, J., L. P. Shoo, C. Graham, and S. E. Williams. 2009. Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? Ecological Modelling **220**:589-594.
Vörösmarty, C. J., C. A. Federer, and A. L. Schloss. 1998. Potential evaporation functions compared on US watersheds: Possible implications for global-scale water balance and terrestrial ecosystem modeling. Journal of Hydrology **207**:147-169.
Vose, R. S., R. L. Schmoyer, P. M. Steurer, and T. C. Peterson. 1992. The Global Historical Climatology Network: Long-term monthly temperature, precipitation, sea level pressure, and station pressure data. CDIAC Communications.
Wang, M. Y., J. E. Overland, and N. A. Bond. 2010. Climate projections for selected large marine ecosystems. Journal of Marine Systems **79**:258-266.
Warren, D. L., R. E. Glor, and M. Turelli. 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. Evolution **62**:2868-2883.
Warren, D. L., R. E. Glor, and M. Turelli. 2010. ENMTools: a toolbox for comparative studies of environmental niche models. Ecography **33**:607-611.
Wiens, J. A. 1989. Spatial scaling in ecology. Functional Ecology **3**:385-397.
Wiens, J. A., D. Stralberg, D. Jongsomjit, C. A. Howell, and M. A. Snyder. 2009. Niches, models, and climate change: assessing the assumptions and uncertainties. Proceedings of the National Academy of Sciences **106**:19729-19736.
Wilby, R., S. Charles, E. Zorita, B. Timbal, P. Whetton, and L. Mearns. 2004. Guidelines for use of climate scenarios developed from statistical downscaling methods. Intergovernmental Panel on Climate Change.
Wood, A. W., L. R. Leung, V. Sridhar, and D. Lettenmaier. 2004. Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. Climatic change **62**:189-216.
Yesson, C., P. W. Brewer, T. Sutton, N. Caithness, J. S. Pahwa, M. Burgess, W. A. Gray, R. J. White, A. C. Jones, and F. A. Bisby. 2007. How global is the global biodiversity information facility? PLoS ONE **2**:e1124.
Zelazowski, P., Y. Malhi, C. Huntingford, S. Sitch, and J. B. Fisher. 2011. Changes in the potential distribution of humid tropical forests on a warmer planet. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **369**:137-160.
Zhang, J. and M. F. Goodchild. 2002. Uncertainty in geographical information. Taylor & Francis, London, UK.
Zimmermann, N. E., T. C. Edwards, C. H. Graham, P. B. Pearman, and J. C. Svenning. 2010. New trends in species distribution modelling. Ecography **33**:985-989.
Zimmermann, N. E., N. G. Yoccoz, T. C. Edwards, E. S. Meier, W. Thuiller, A. Guisan, D. R. Schmatz, and P. B. Pearman. 2009. Climatic extremes improve predictions of spatial patterns of tree species. Proceedings of the National Academy of Sciences of the United States of America **106**:19723-19728.
Zweig, M. H. and G. Campbell. 1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical Chemistry **39**:561-577.

Appendix A

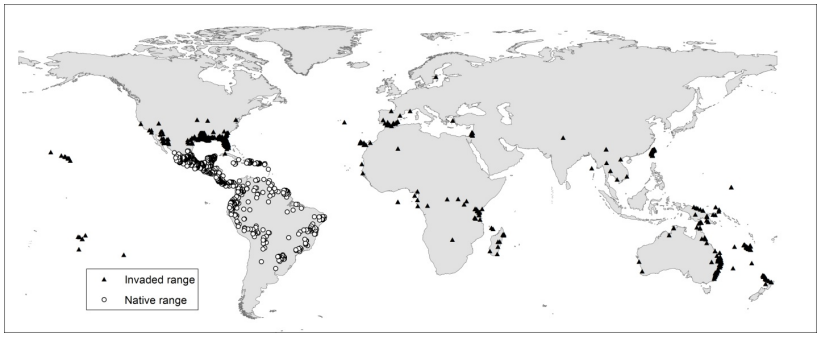

Figure S1. *Lantana camara* georeferenced occurrences.

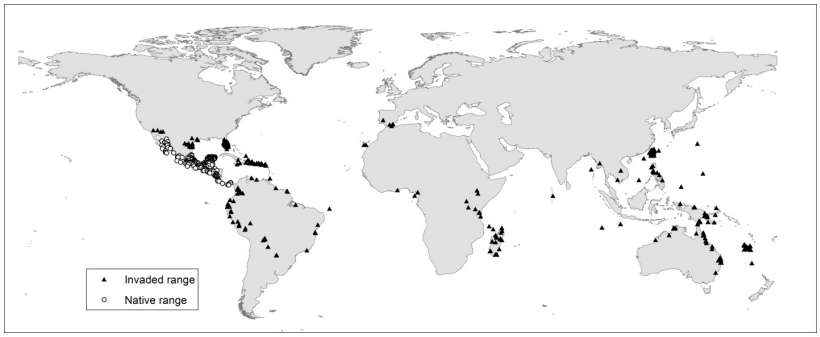

Figure S2. *Leucaena leucocephala* georeferenced occurrences.

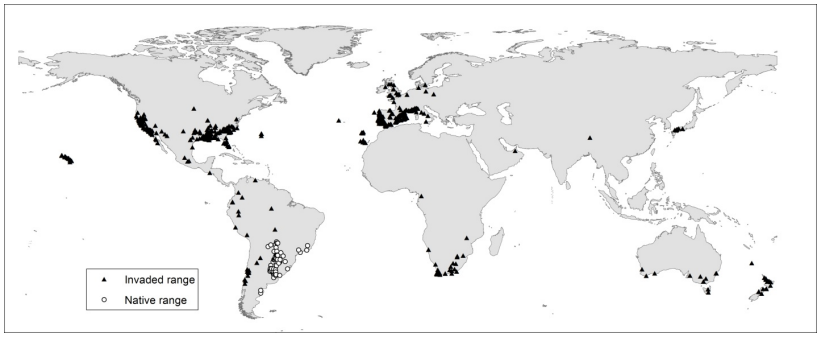

Figure S3. *Linepithema humile* georeferenced occurrences.

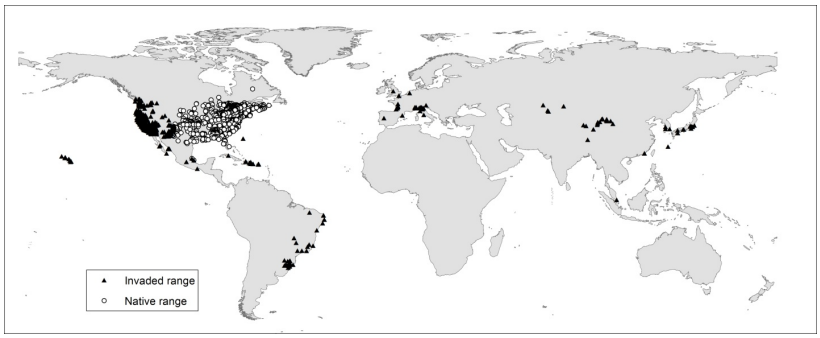

Figure S4. *Lithobates catesbeianus* georeferenced occurrences.

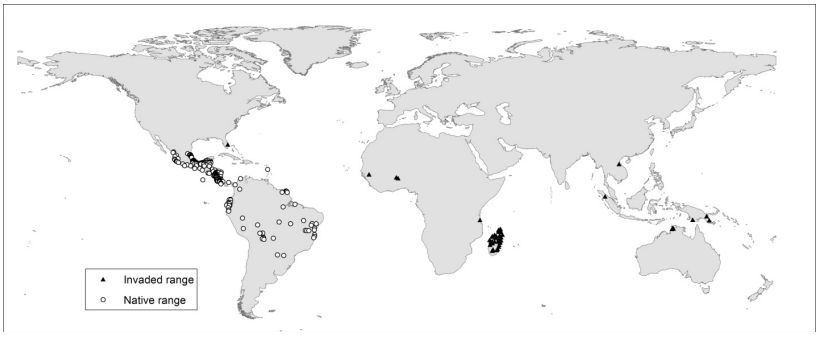

Figure S5. *Mimosa pigra* georeferenced occurrences.

Figure S6. *Rhinella marina* georeferenced occurrences.

Figure S7. *Sphagneticola trilobata* georeferenced occurrences.

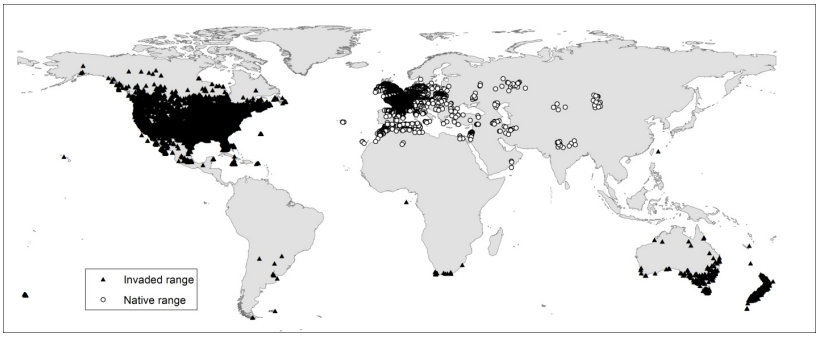

Figure S8. *Sturnus vulgaris* georeferenced occurrences.

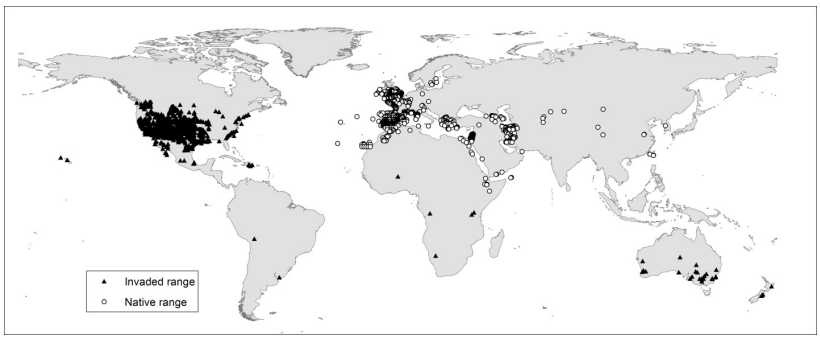

Figure S9. *Tamarix ramosissima* georeferenced occurrences.

Figure S10. *Ulex europaeus* georeferenced occurrences.

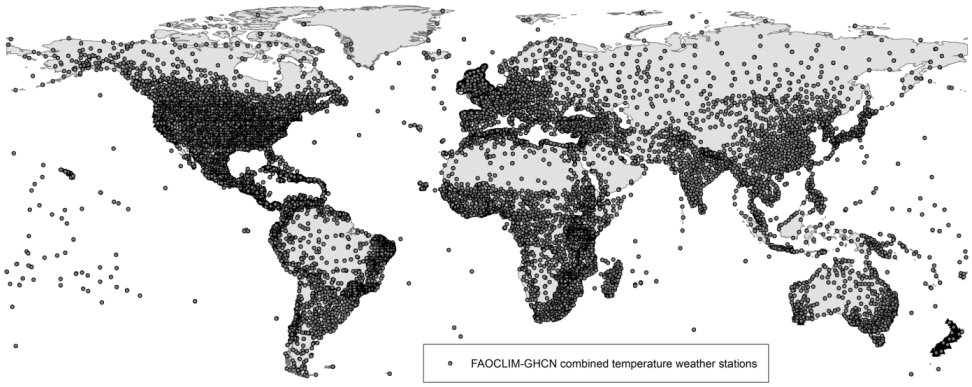

Figure S11. FAOCLIM and GHCN temperature weather station locality maps.

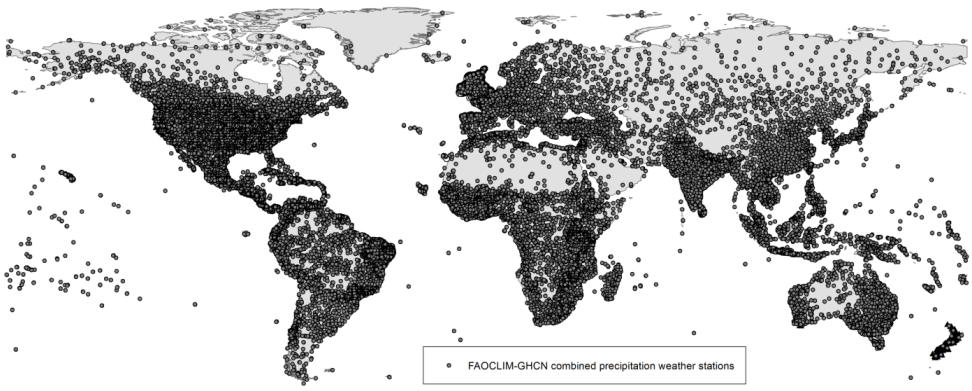

Figure S12. FAOCLIM and GHCN precipitation weather station locality maps.

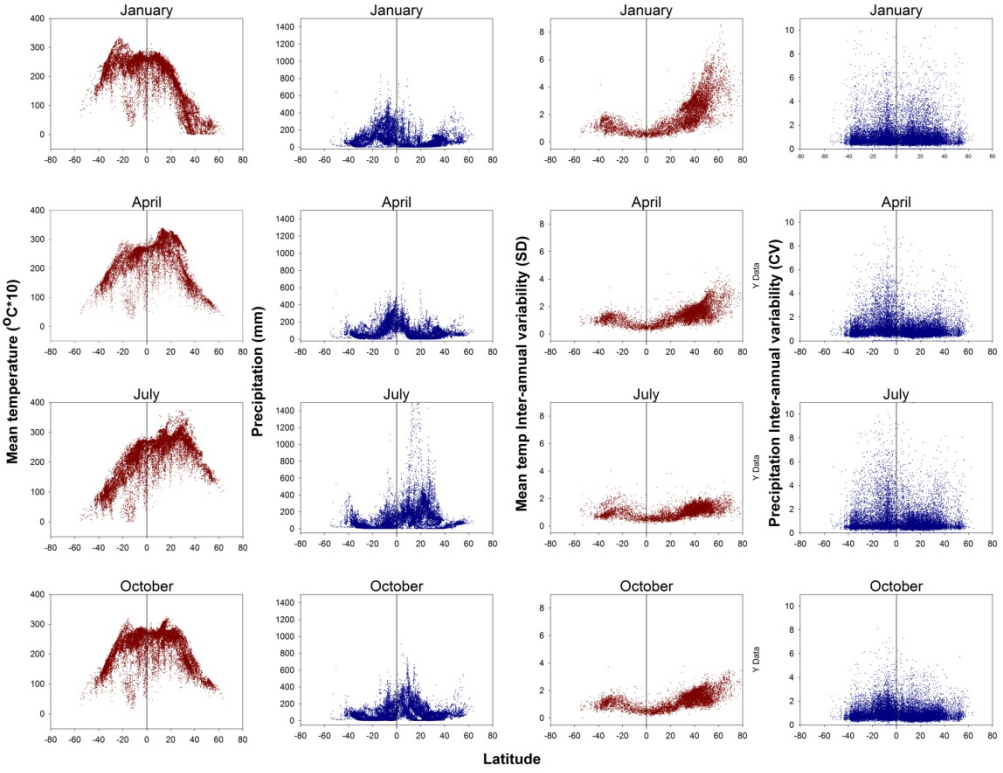

Figure S13. Global reverse seasonality evaluation. We plotted values of weather stations against latitude for four purposely selected months that represent the peak of each season in the Northern hemisphere. Red represent stations that record temperature and blue represent stations that record precipitation. The **first and second** column represent mean monthly temperature and monthly total precipitation and how the change along the latitudinal gradient. The reverse pattern is clearly depicted by mean temperature where January almost mirrors July. Precipitation also shows a clear reverse pattern between January and July. The **third and fourth** columns represent inter-annual variability

values. Red depicts standard deviation for temperature recording weather stations and blue depicts coefficient of variation for precipitation recording stations. Note that the there is no reverse pattern in the third and fourth column which represent ClimVar.

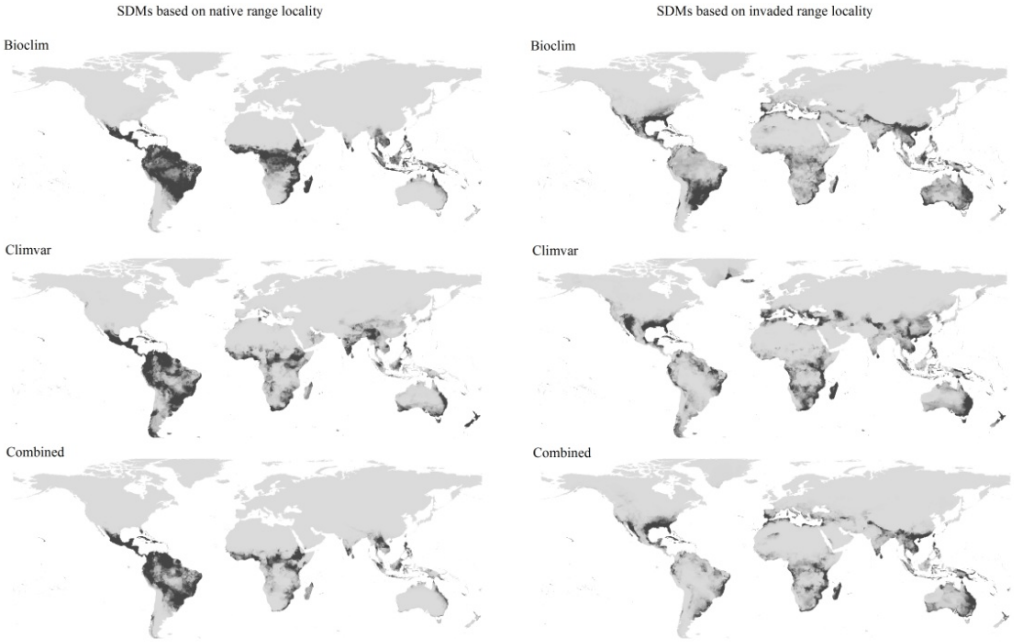

Figure S14. *Lantana camara* Maxent logistic model outputs. Left: models based on the native range localities. Right: models based on the invaded range localities.

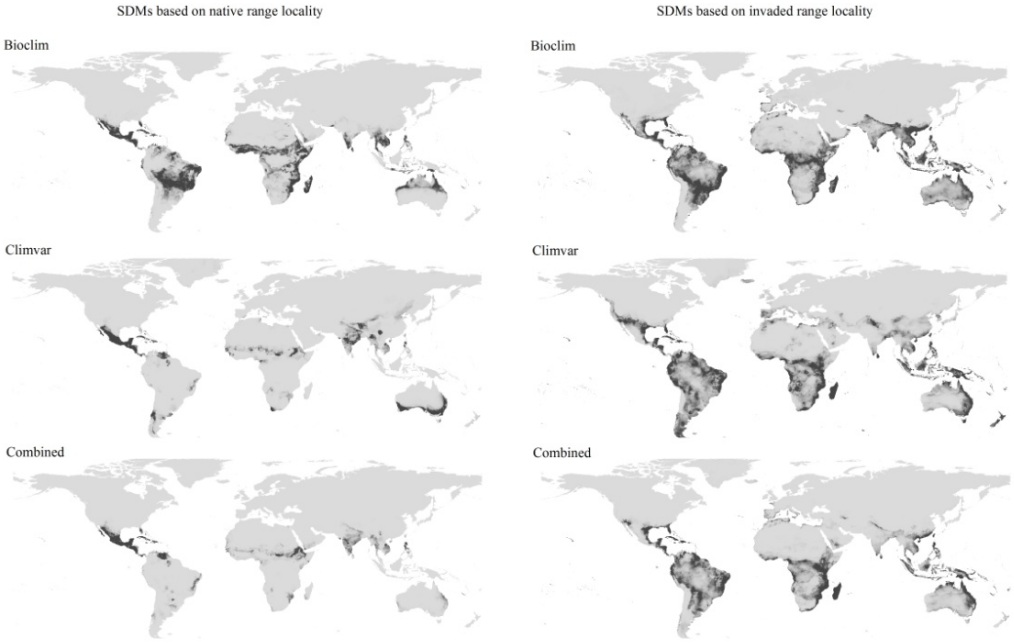

Figure S15. *Leucaena leuocephala* Maxent logistic model outputs. Left: models based on the native range localities. Right: models based on the invaded range localities.

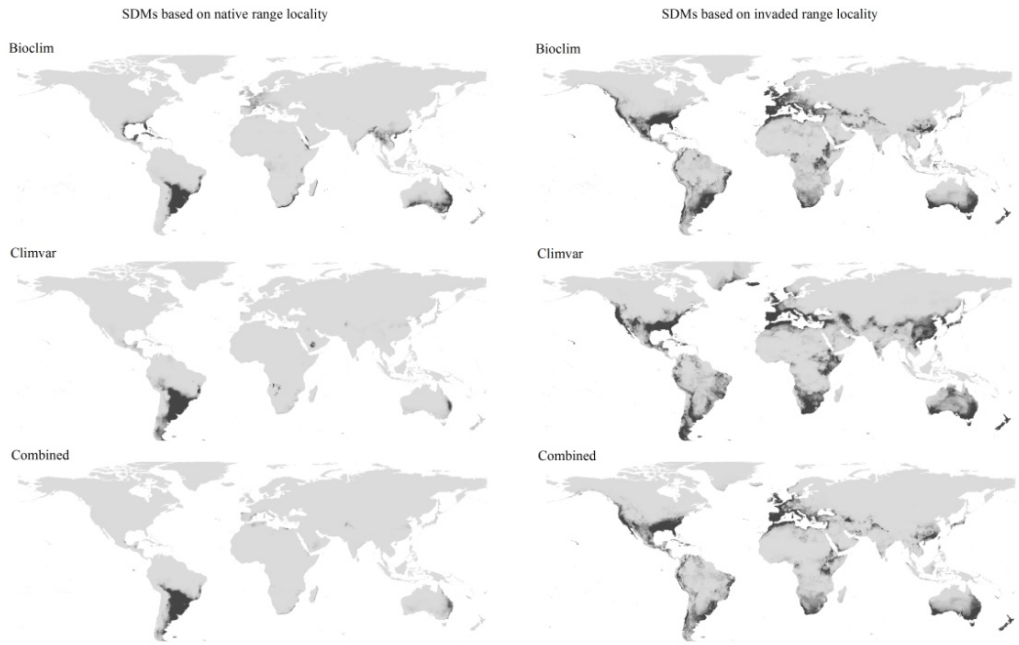

Figure S16. *Linepithema humile* Maxent logistic model outputs. Left: models based on the native range localities. Right: models based on the invaded range localities.

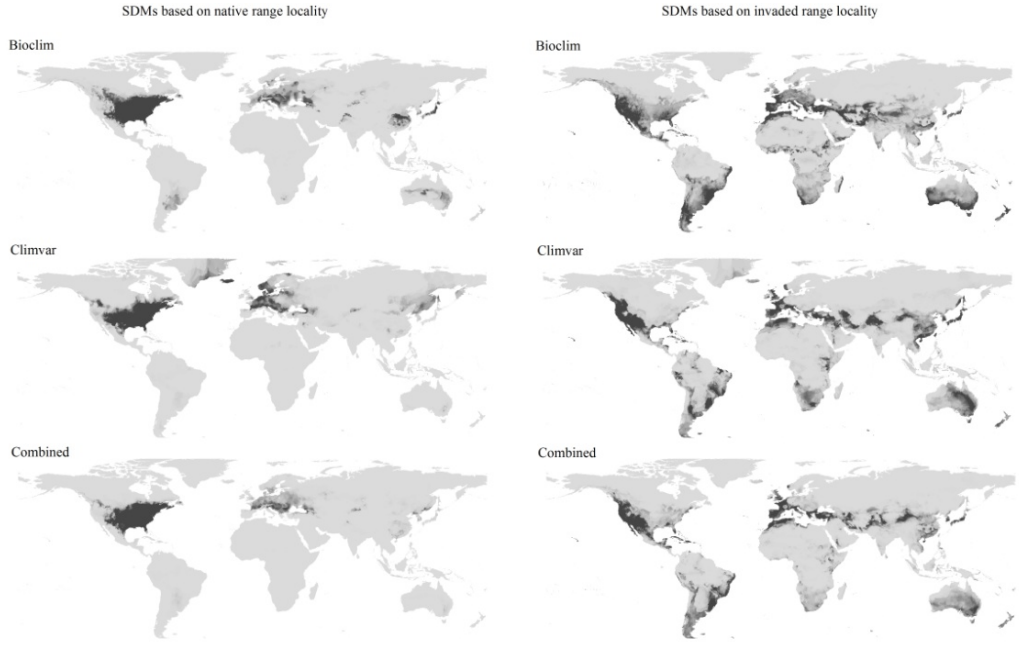

Figure S17. *Lithobates catesbeianus* Maxent logistic model outputs. Left: models based on the native range localities. Right: models based on the invaded range localities.

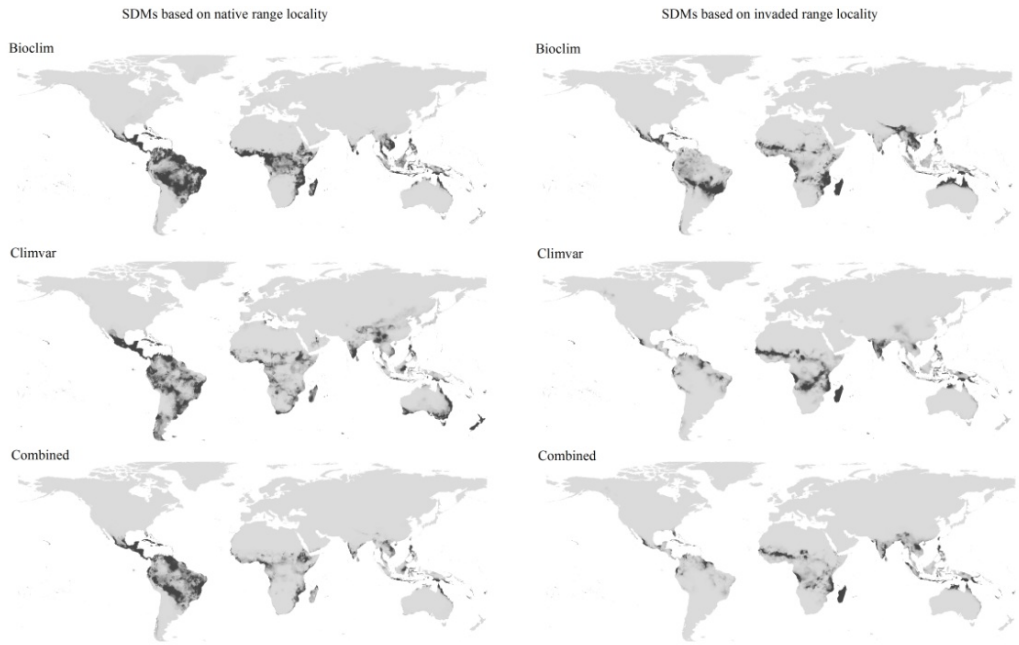

Figure S18. *Mimosa pigra* Maxent logistic model outputs. Left: models based on the native range localities. Right: models based on the invaded range localities.

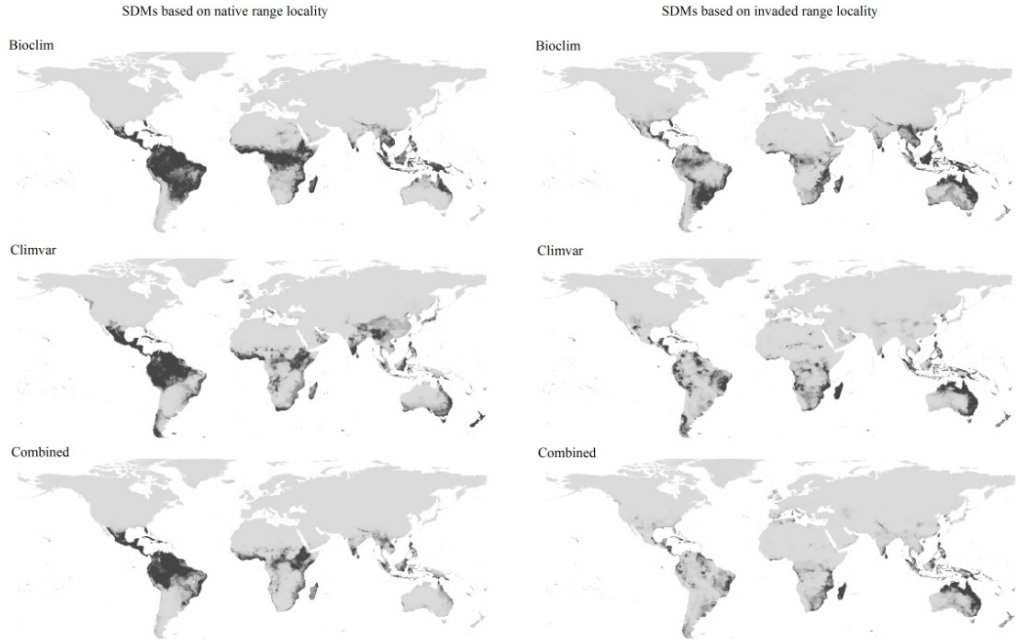

Figure S19. *Rhinella marina* Maxent logistic model outputs. Left: models based on the native range localities. Right: models based on the invaded range localities.

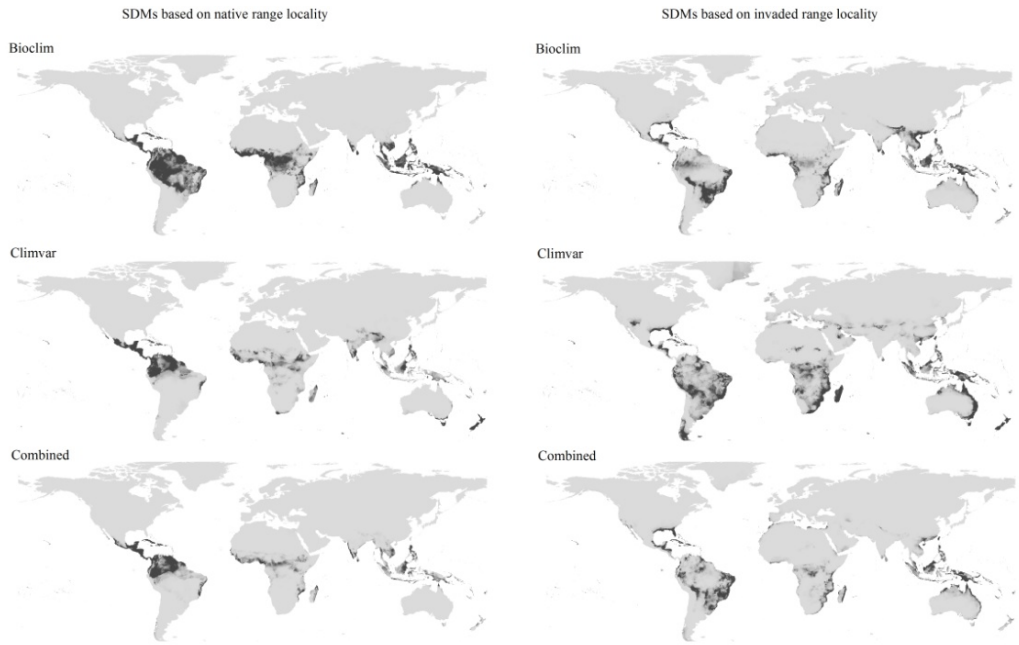

Figure S20. *Spagneticola trilobata* Maxent logistic model outputs. Left: models based on the native range localities. Right: models based on the invaded range localities.

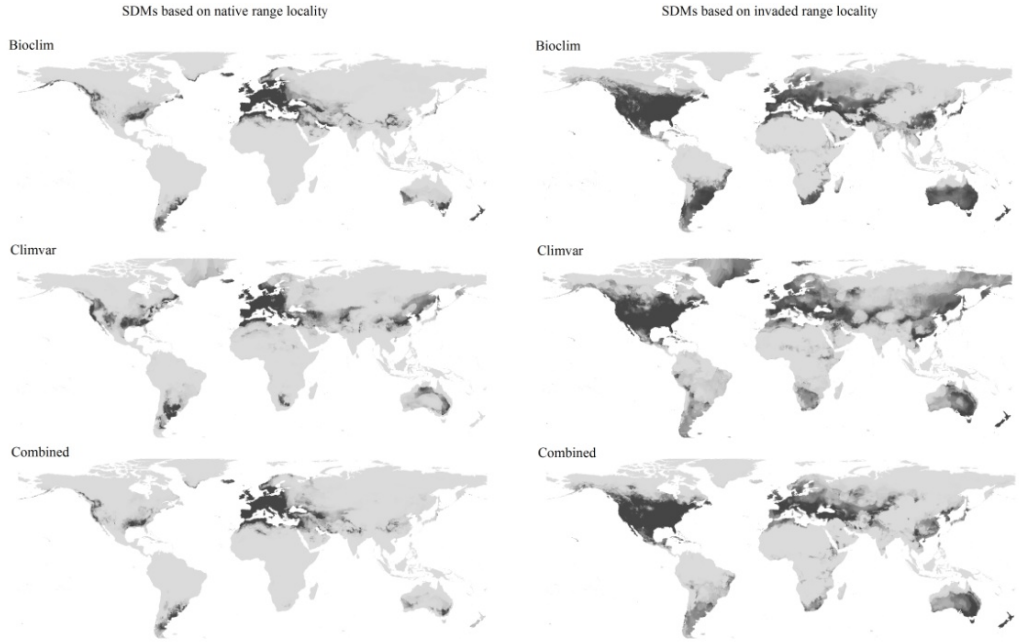

Figure S21. *Sturnus vulgaris* Maxent logistic model outputs. Left: models based on the native range localities. Right: models based on the invaded range localities.

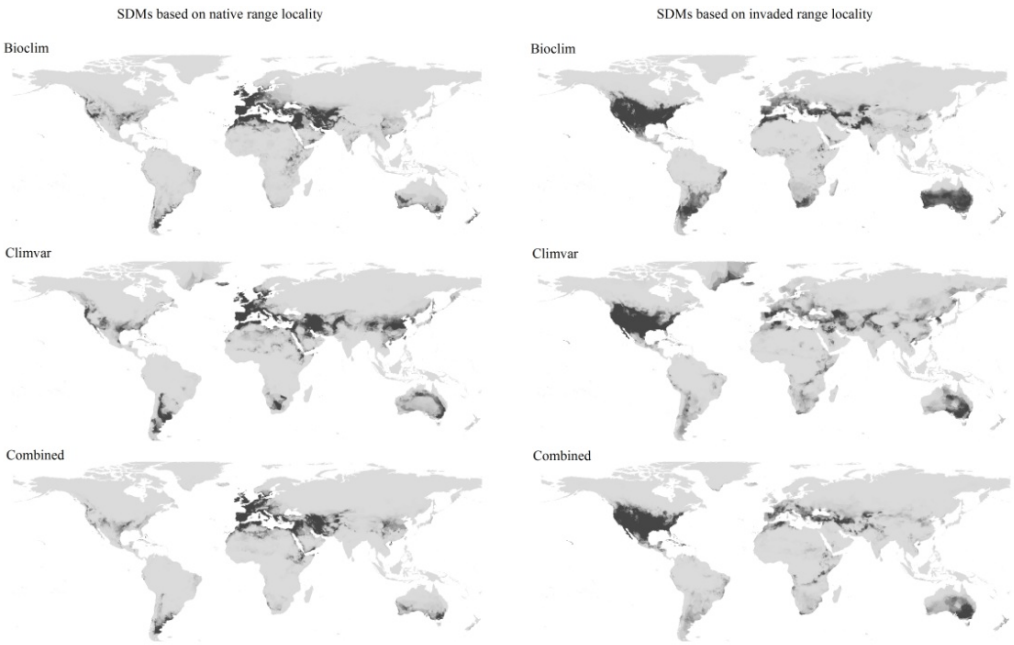

Figure S22. *Tamarix ramosissima* Maxent logistic model outputs. Left: models based on the native range localities. Right: models based on the invaded range localities.

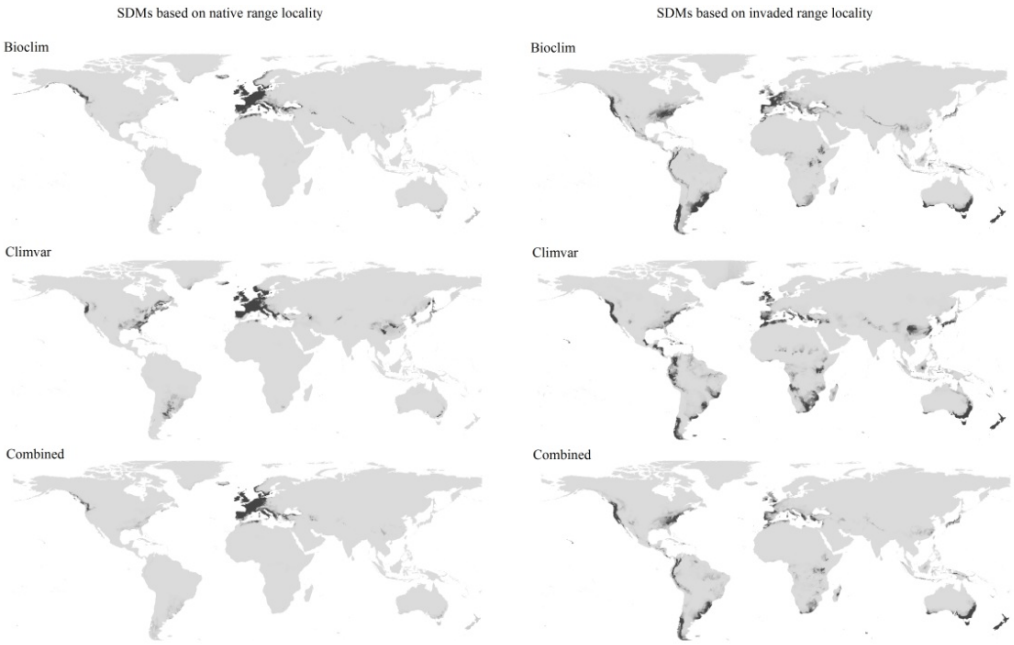

Figure S23. *Ulex europaeus* Maxent logistic model outputs. Left: models based on the native range localities. Right: models based on the invaded range localities.

	Native range						Invaded range					
	Treatment			Contrast			Treatment			Contrast		
Species	AUC1	AUC2	AUC3	1/2	1/3	2/3	AUC1	AUC2	AUC3	1/2	1/3	2/3
Lantana camara	0.967	0.961	0.974	**	**	**	0.975	0.961	0.975	**		**
Leucaena leucocephala	0.986	0.960	0.986	**		**	0.965	0.960	0.973	**	**	**
Linepithema humile	0.989	0.981	0.992	**	**	**	0.972	0.957	0.975	**	**	**
Lithobates catesbeianus	0.974	0.961	0.976	**	**	**	0.968	0.962	0.977	**	**	**
Mimosa pigra	0.984	0.972	0.984	**		**	0.991	0.979	0.991	**		**
Rhinella marina	0.967	0.960	0.973	**	**	**	0.977	0.970	0.986	**	**	**
Sphagneticola trilobata	0.987	0.978	0.987	**		**	0.985	0.971	0.985	**		**
Sturnus vulgaris	0.975	0.957	0.975	**		**	0.974	0.959	0.982	**	**	**
Tamarix ramosissima	0.966	0.969	0.973	**	**	**	0.965	0.961	0.973	**	**	**
Ulex europaeus	0.973	0.968	0.978	**	**	**	0.983	0.977	0.988	**	**	**

**<0.05

Table S1. Native and Invaded range SDM performance, summary of the models and comparison tests. AUC1: Bioclim environmental layer configuration. AUC2: ClimVar environmental layer configuration. AUC3: combined environmental layer configuration. The significance of the difference between models in an ANOVA with a Fisher's least significant difference procedure is reported for the following comparisons: 1/2: Bioclim over ClimVar; 1/3: Bioclim over Bioclim-ClimVar; 2/3: ClimVar over Bioclim-ClimVar.

Species	**Treatment**			**Contrast**		
	CONF1	CONF2	CONF3	1/2	1/3	2/3
Lantana camara	0.5664	0.5405	0.4922	**	**	**
Leucaena leucocephala	0.5825	0.4418	0.4424	**	**	
Linepithema humile	0.5075	0.4051	0.4125	**	**	**
Lithobates catesbeianus	0.4504	0.4119	0.4038	**	**	**
Mimosa pigra	0.6102	0.5063	0.4729	**	**	**
Rhinella marina	0.6395	0.5729	0.5194	**	**	**
Sphagneticola trilobata	0.6007	0.4881	0.5041	**	**	**
Sturnus vulgaris	0.5477	0.5734	0.4892	**	**	**
Tamarix ramosissima	0.5153	0.4935	0.4534	**	**	**
Ulex europaeus	0.4606	0.4147	0.4321	**	**	**

**<0.05

Table S2. Niche similarity between native and invaded range, summary of the models and comparison tests. CONF1: Bioclim environmental layer configuration. CONF2: ClimVar environmental layer configuration. CONF3: combined environmental layer configuration. The significance of the difference between models in an ANOVA with a Fisher's least significant difference procedure is reported for the following comparisons:

1/2: Bioclim over ClimVar; 1/3: Bioclim over Bioclim-ClimVar; 2/3: ClimVar over Bioclim-ClimVar.

APPENDIX B

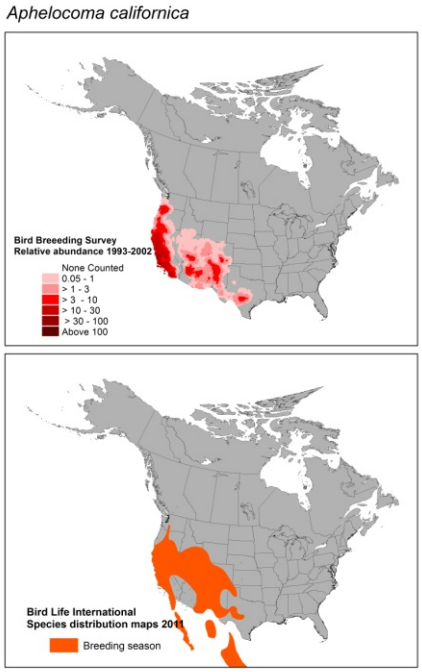

Figure S24. *Aphlecoma californica* distribution, Top: Bird breeding Survey data; bottom: Bird Life International-NatureServe data.

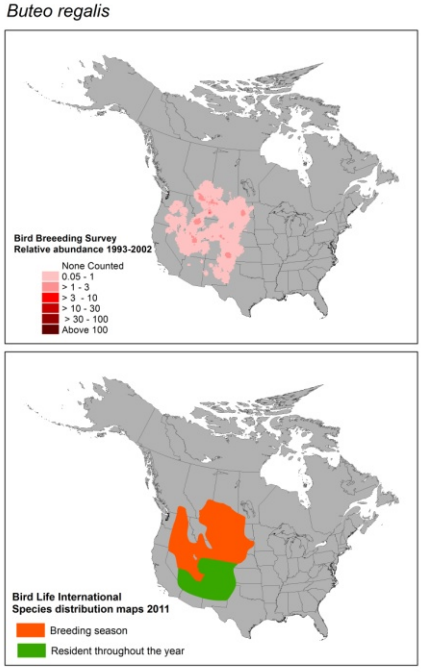

Figure S25. *Buteo regalis* distribution, Top: Bird breeding Survey data; bottom: Bird Life International-NatureServe data.

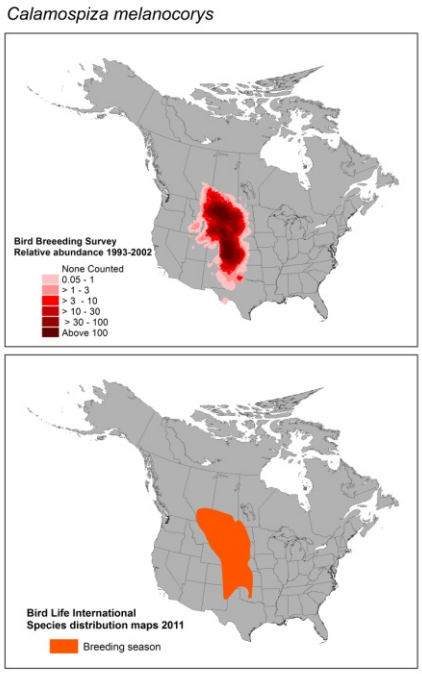

Figure S26. *Calamospiza melanocorys* distribution, Top: Bird breeding Survey data; bottom: Bird Life International-NatureServe data.

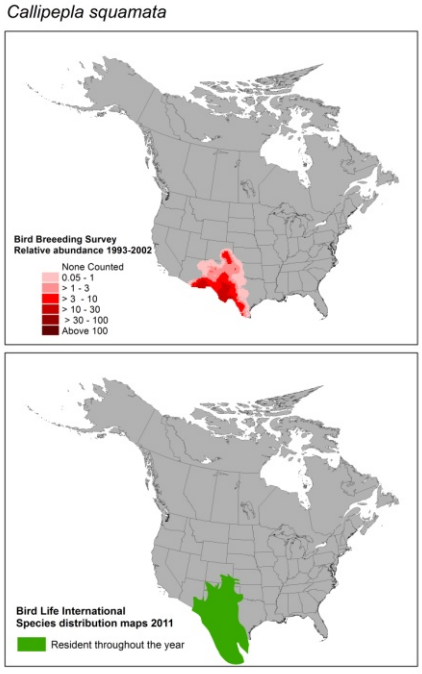

Figure S27. *Callipepla squamata* distribution, Top: Bird breeding Survey data; bottom: Bird Life International-NatureServe data.

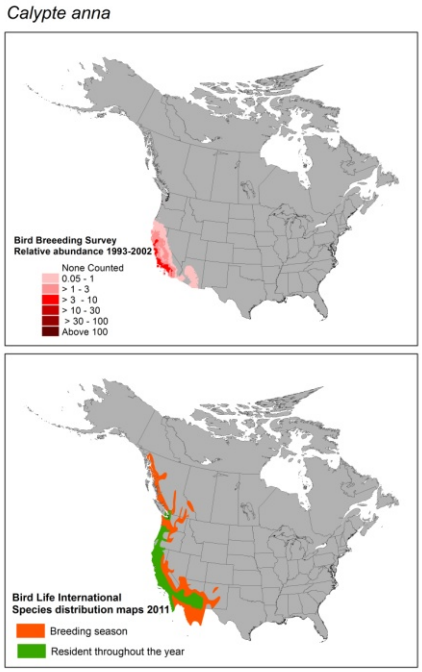

Figure S28. *Calypte anna* distribution, Top: Bird breeding Survey data; bottom: Bird Life International-NatureServe data.

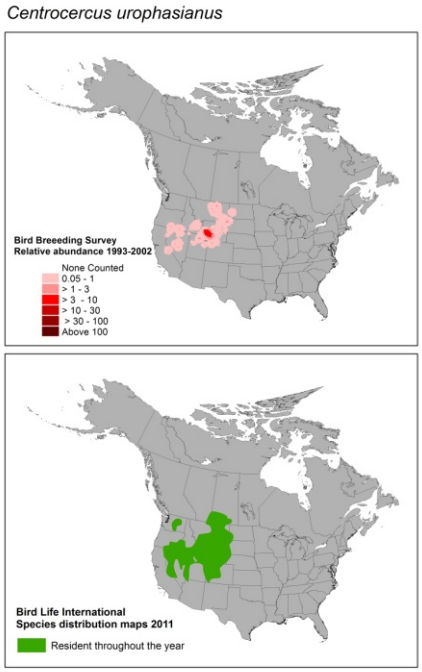

Figure S29. *Centrocercus urophasianus* distribution, Top: Bird breeding Survey data; bottom: Bird Life International-NatureServe data.

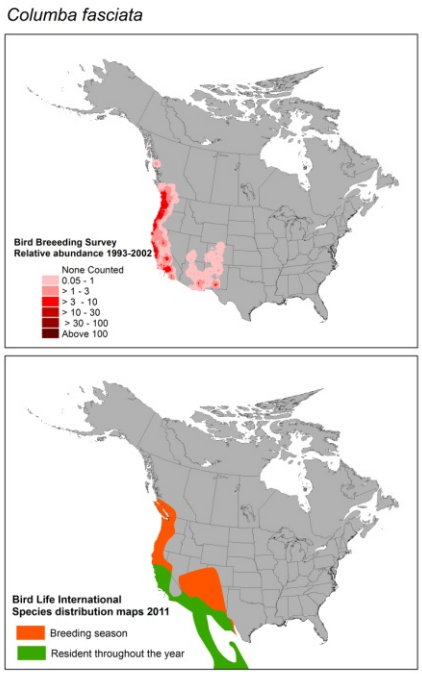

Figure S30. *Columba fasciata* distribution, Top: Bird breeding Survey data; bottom: Bird Life International-NatureServe data.

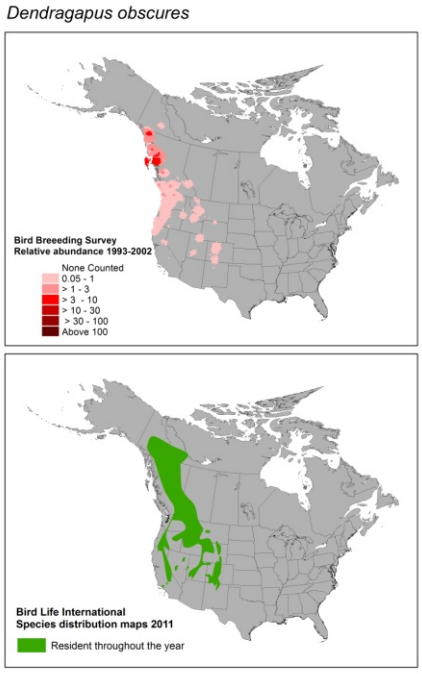

Figure S31. *Dendragapus obscures* distribution, Top: Bird breeding Survey data; bottom: Bird Life International-NatureServe data.

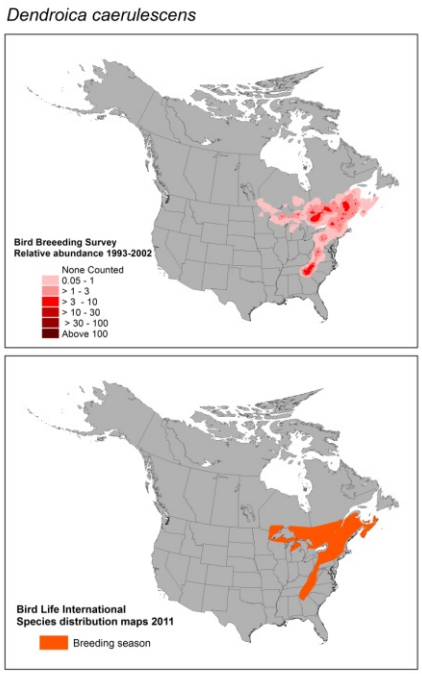

Figure S32. *Dendroica caerulescens* distribution, Top: Bird breeding Survey data; bottom: Bird Life International-NatureServe data.

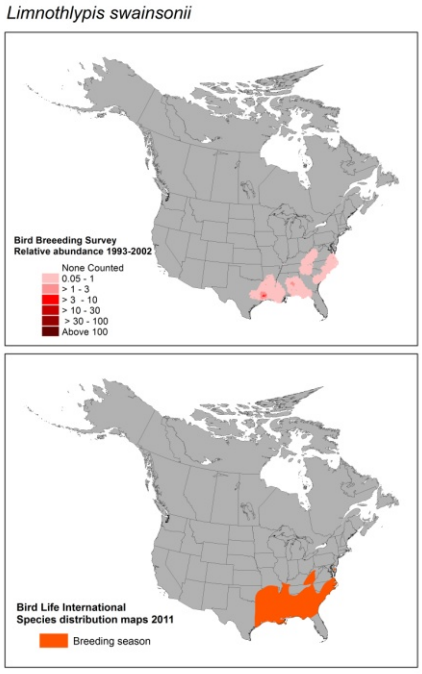

Figure S33. *Limnothlypis swainsonii* distribution, Top: Bird breeding Survey data; bottom: Bird Life International-NatureServe data.

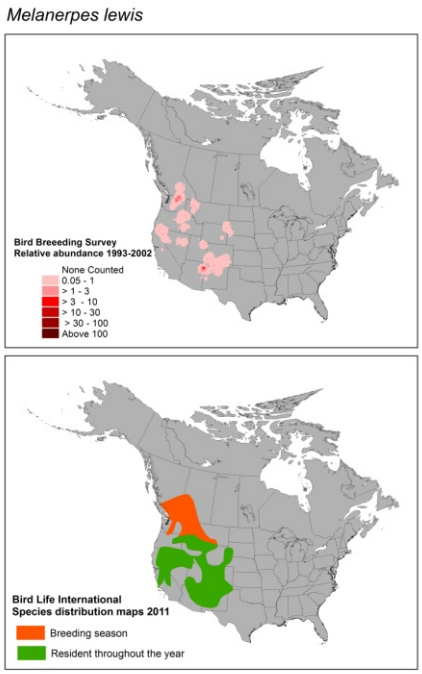

Figure S34. *Melanerpes lewis* distribution, Top: Bird breeding Survey data; bottom: Bird Life International-NatureServe data.

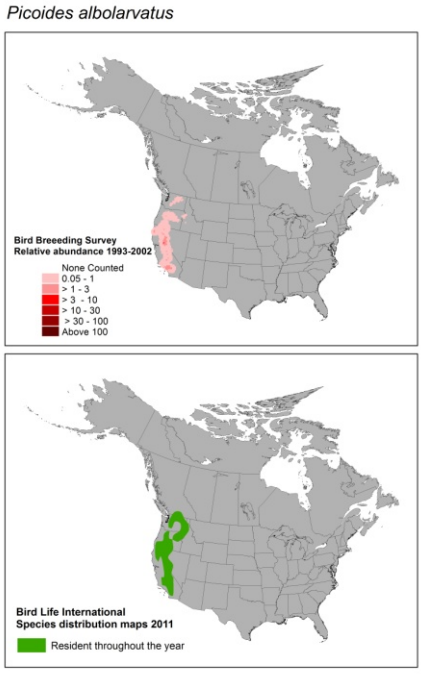

Figure S35. *Picoides albolarvatus* distribution, Top: Bird breeding Survey data; bottom: Bird Life International-NatureServe data.

Picoides borealis

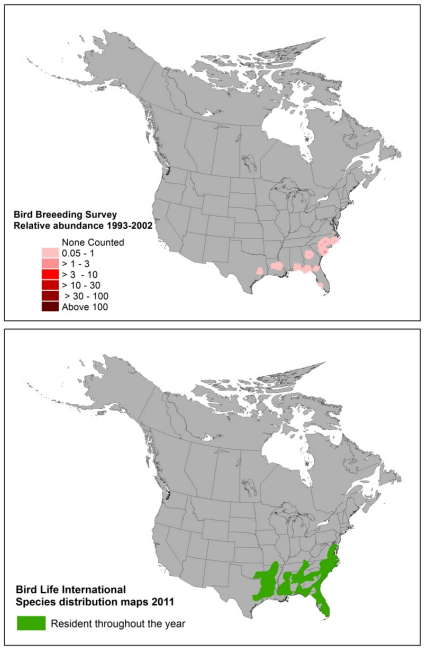

Figure S36. *Picoides borealis* distribution, Top: Bird breeding Survey data; bottom: Bird Life International-NatureServe data.

Picoides nuttallii

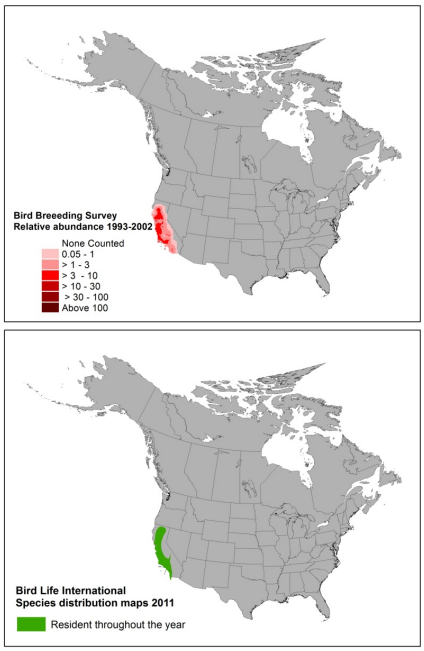

Figure S37. *Picoides nuttalli* distribution, Top: Bird breeding Survey data; bottom: Bird Life International-NatureServe data.

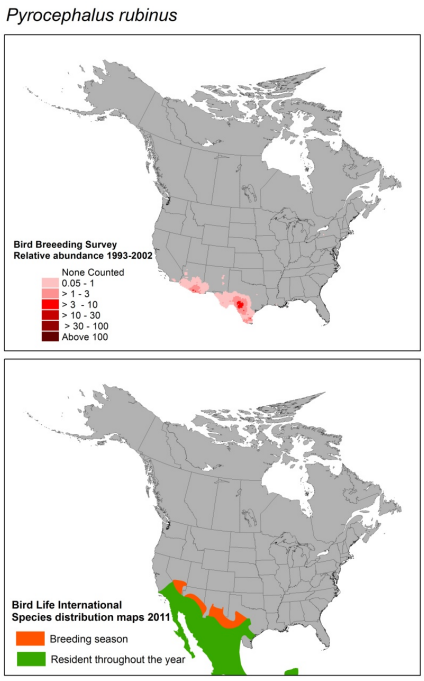

Figure S38. *Pyrocephalus rubinus* distribution, Top: Bird breeding Survey data; bottom: Bird Life International-NatureServe data.

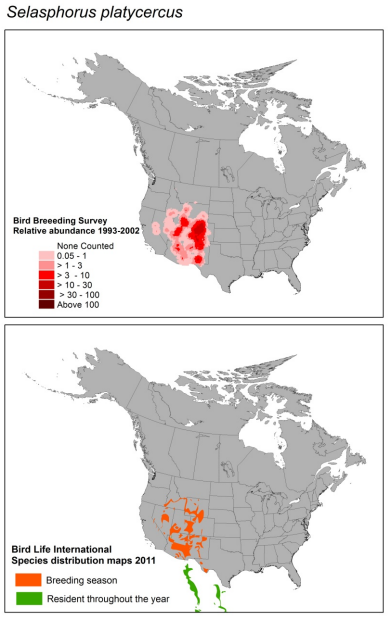

Figure S39. *Selasphorus platycercus* distribution, Top: Bird breeding Survey data; bottom: Bird Life International-NatureServe data.

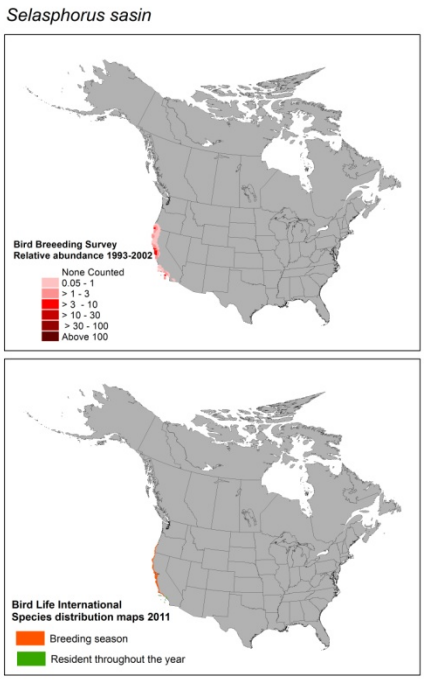

Figure S40. *Selasphorus sasin* distribution, Top: Bird breeding Survey data; bottom: Bird Life International-NatureServe data.

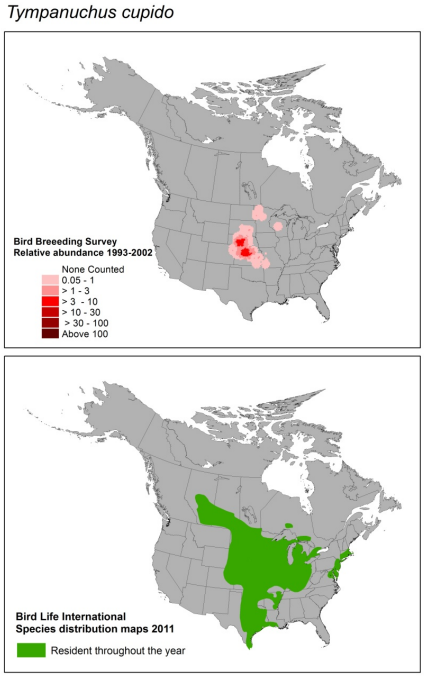

Figure S41. *Tympanuchus cupido* distribution, Top: Bird breeding Survey data; bottom: Bird Life International-NatureServe data.

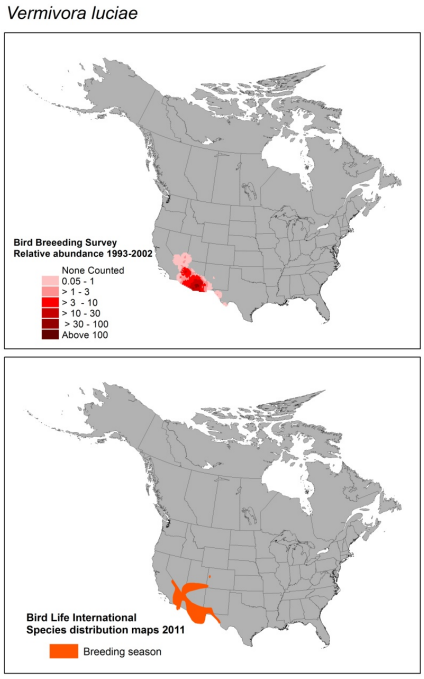

Figure S42. *Vermivora luciae* distribution, Top: Bird breeding Survey data; bottom: Bird Life International-NatureServe data.

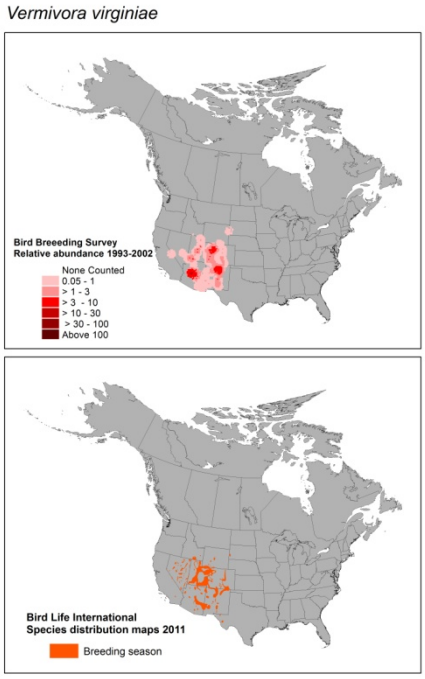

Figure S43. *Vermivora virginiae* distribution, Top: Bird breeding Survey data; bottom: Bird Life International-NatureServe data.

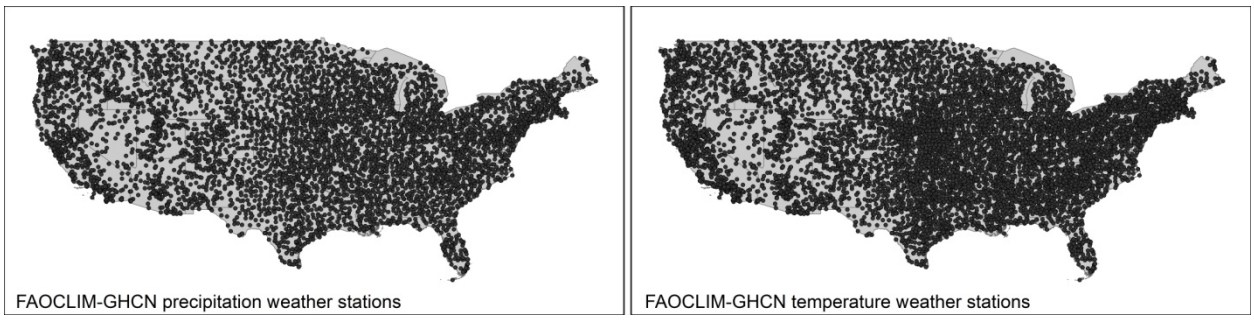

Figure S44. Precipitation (left) and temperature (right) weather stations combined from FAOCLIM-2 and GHCN.

APPENDIX C

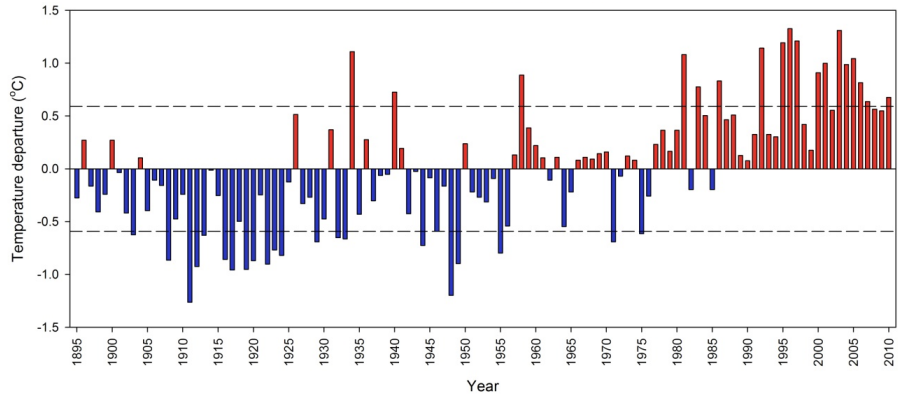

Figure S45. Minimum annual temperature anomalies calculated from 1895 to 2010 base period.

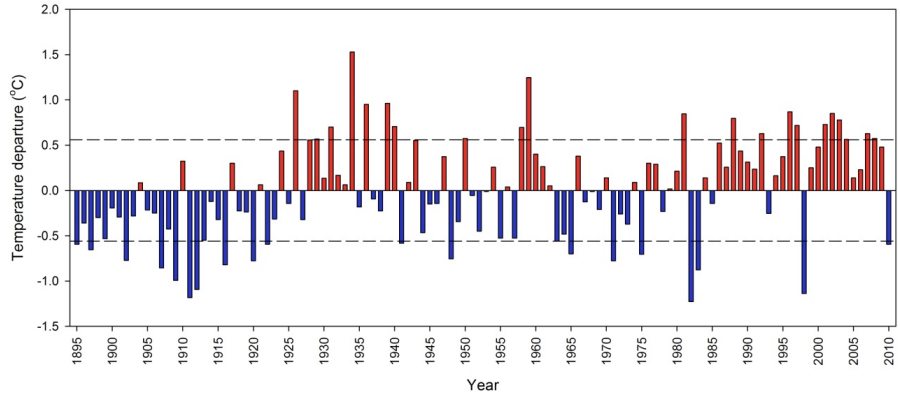

Figure S46. Maximum annual temperature anomalies calculated from 1895 to 2010 base period.

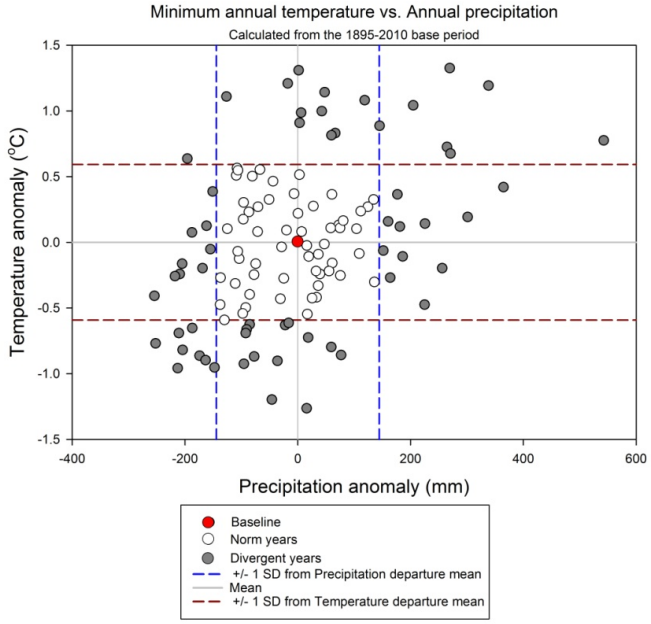

Figure S47. Assignment of individual years to climate scenarios. Each white or gray circle represents minimum annual temperature and total precipitation anomalies for an individual year between1895 and 2010.

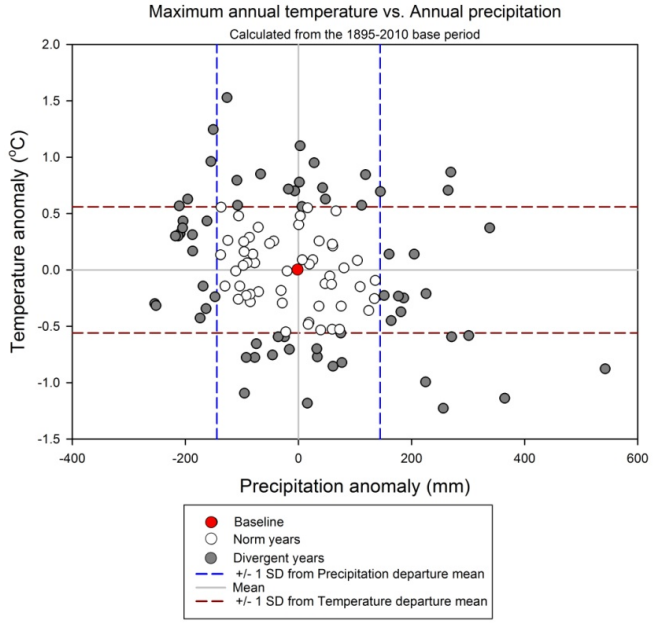

Figure S48. Assignment of individual years to climate scenarios. Each white or gray circle represents maximum annual temperature and total precipitation anomalies for an individual year between1895 and 2010.

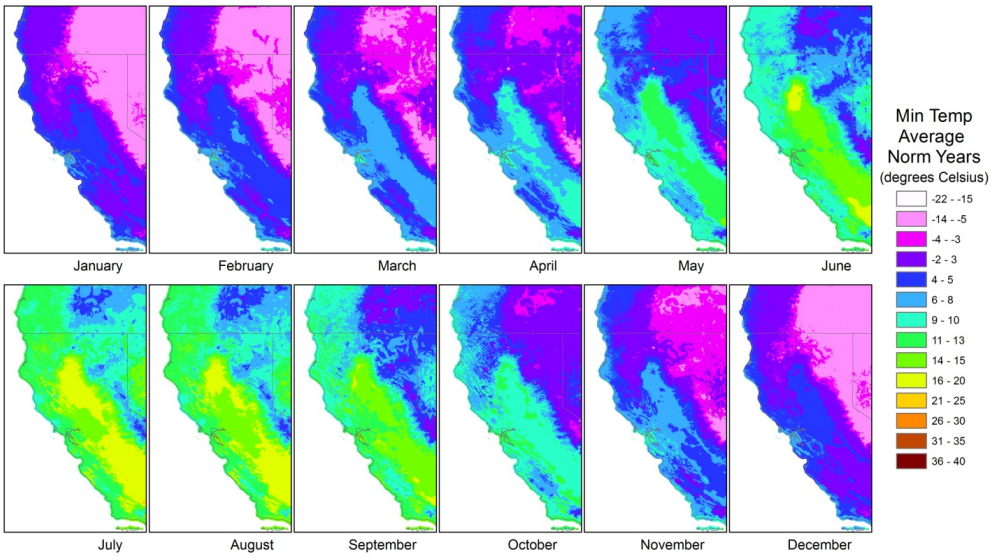

Figure S49. PRISM monthly minimum temperatures baseline.

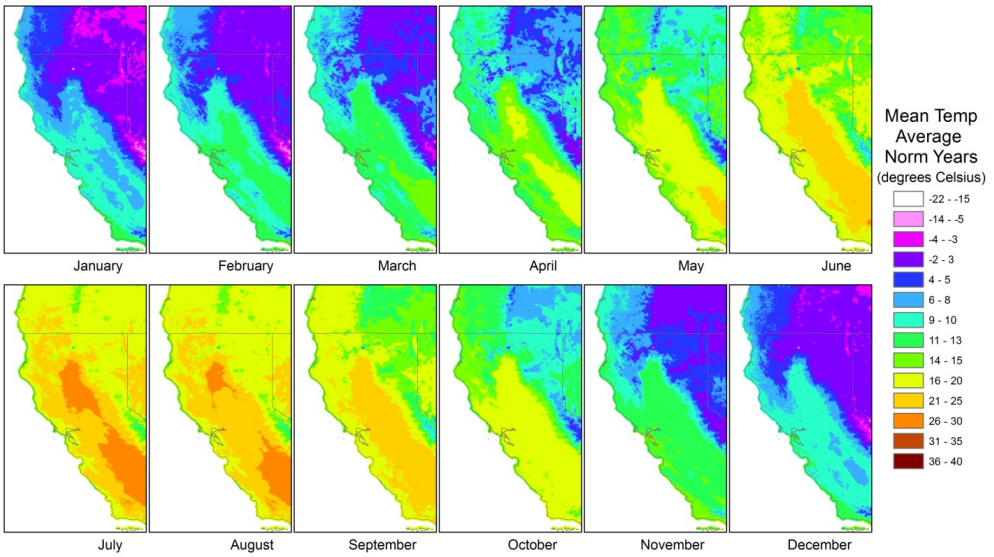

Figure S50. PRISM monthly mean temperatures baseline.

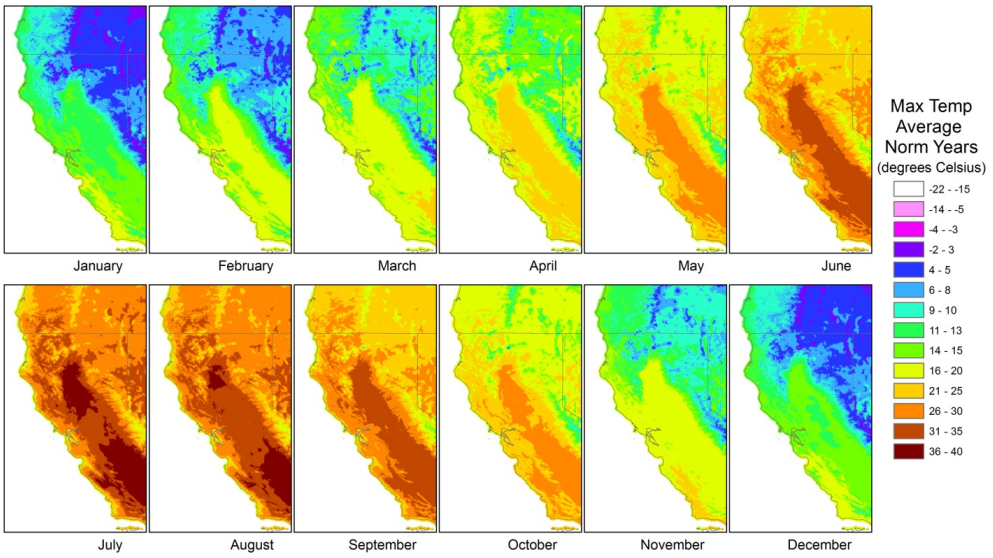

Figure S51. PRISM monthly maximum temperatures baseline.

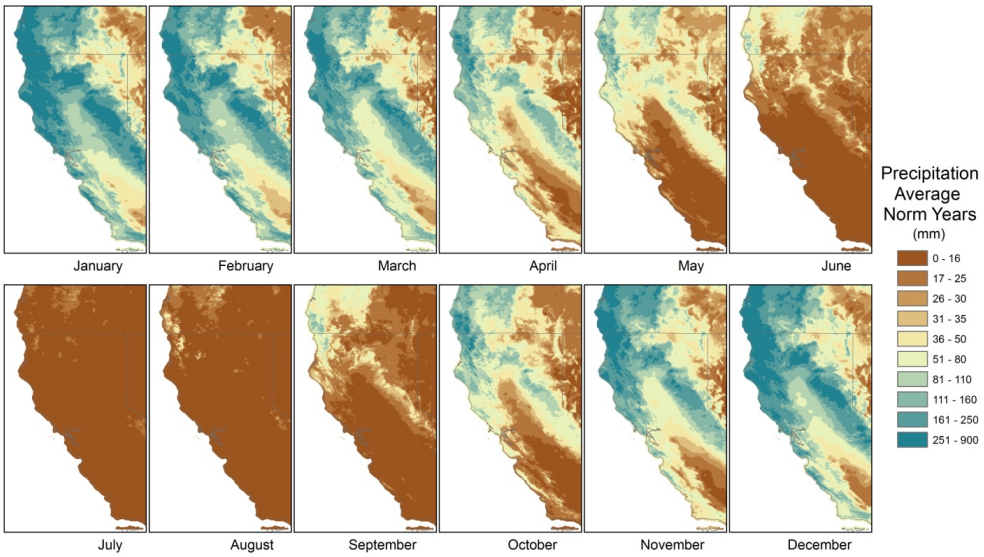

Figure S52. PRISM monthly total precipitation baseline.

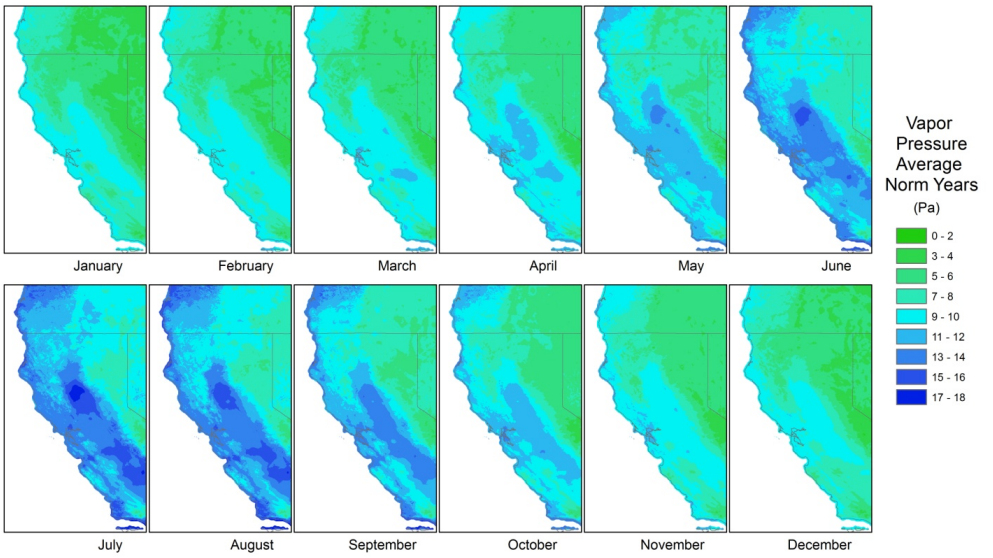

Figure S53. PRISM monthly vapor pressure baseline.

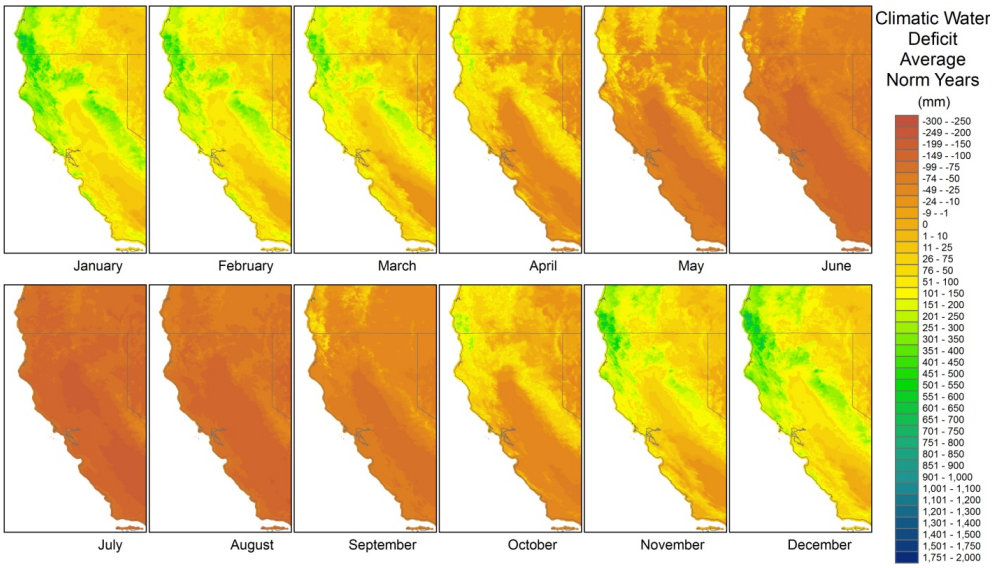

Figure S54. Climatic water deficit derived from PRISM baseline.

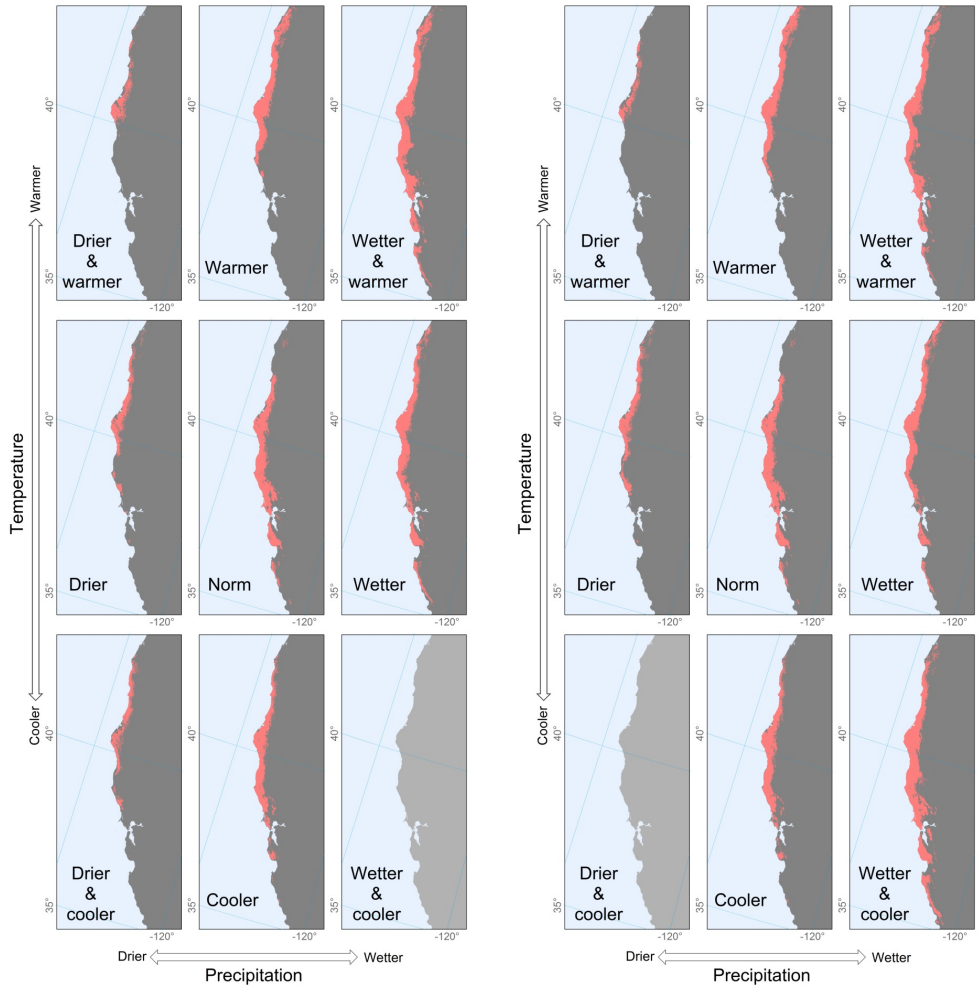

Figure S55. Predicted suitable habitat for the eight scenarios we developed. Left, based on years selected from the minimum temperature and total annual precipitation anomalies. Right, based on years selected from the maximum temperature and annual precipitation anomalies.

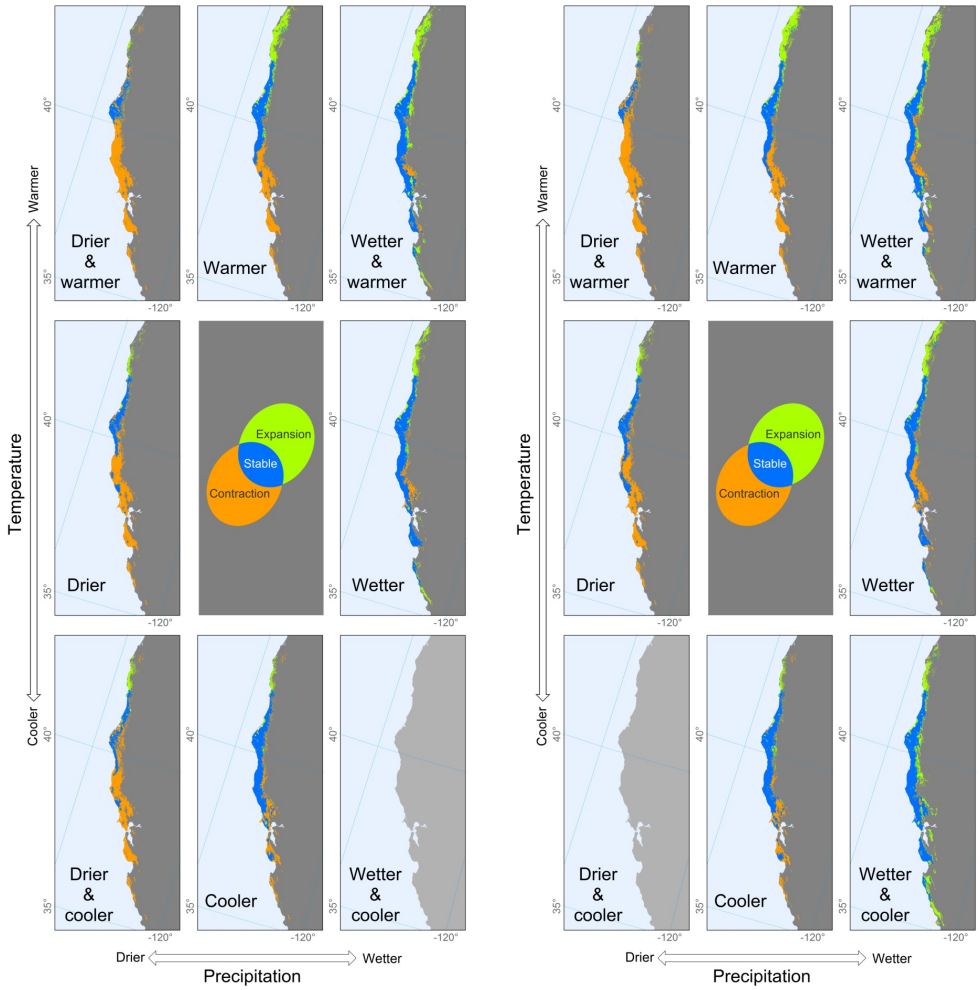

Figure S56. Synthetic generalization of the predicted expansion, contraction and stability for the eight scenarios we developed. Left, based on years selected from the minimum temperature and total annual precipitation anomalies. Right, based on years selected from the maximum temperature and annual precipitation anomalies.

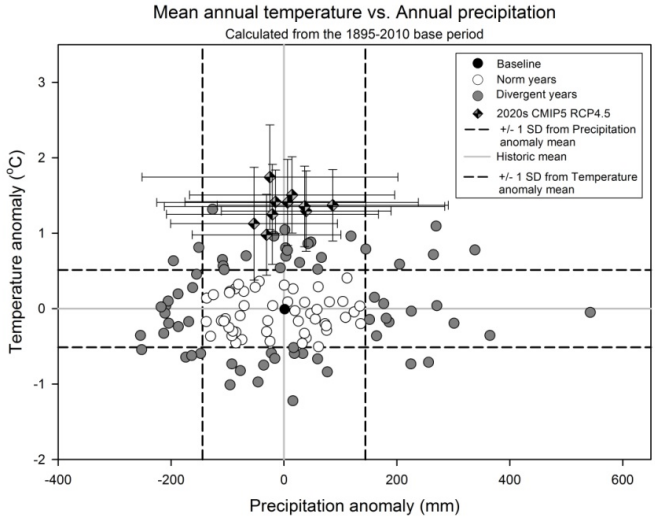

Figure S57. Multi-model mean annual temperature and precipitation anomalies for California projected for the 2020s in the 21st century compared to historical annual temperature and precipitation anomalies (circles as in Figure 10).

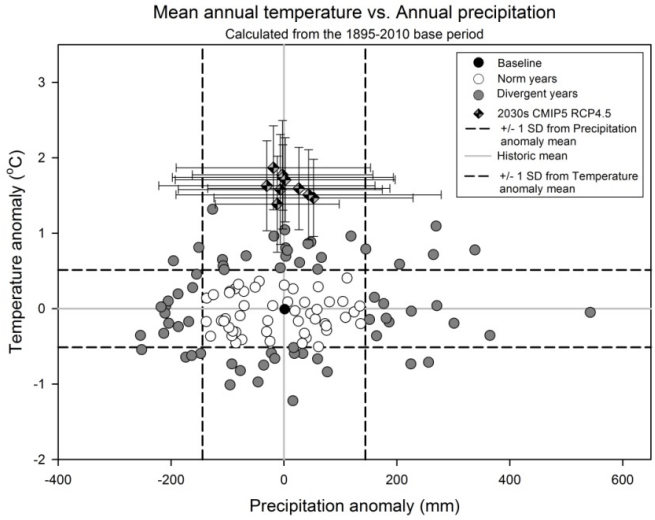

Figure S58. Multi-model mean annual temperature and precipitation anomalies for California projected for the 2030s in the 21st century compared to historical annual temperature and precipitation anomalies (circles as in Figure 10).

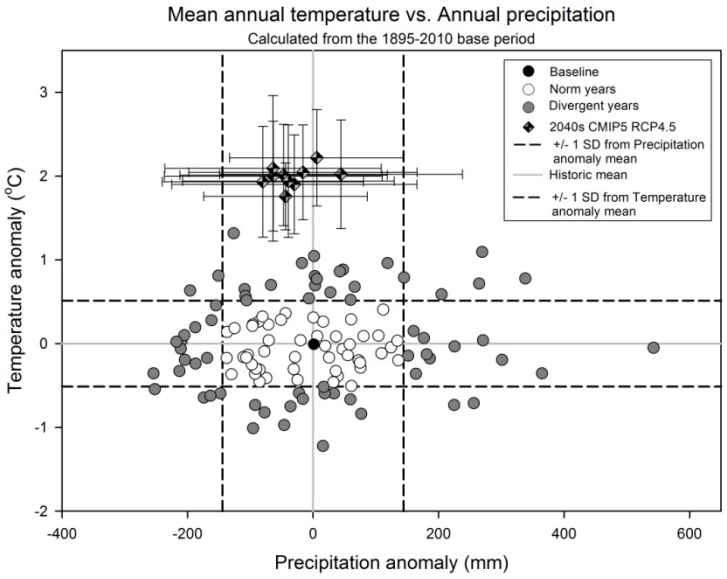

Figure S59. Multi-model mean annual temperature and precipitation anomalies for California projected for the 2040s in the 21st century compared to historical annual temperature and precipitation anomalies (circles as in Figure 10).

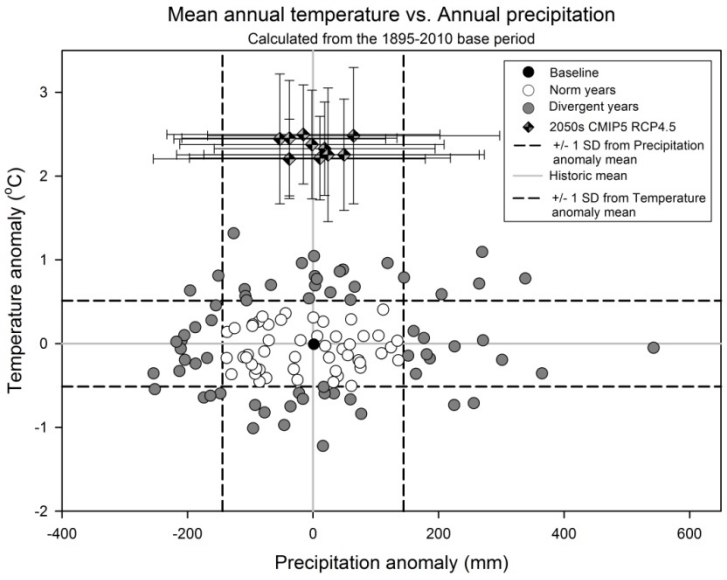

Figure S60. Multi-model mean annual temperature and precipitation anomalies for California projected for the 2050s in the 21st century compared to historical annual temperature and precipitation anomalies (circles as in Figure 10).