UNIVERSITY OF CALIFORNIA, MERCED

UNDERSTATING CLIMATE CHANGE AND THE IMPACT ON BIODIVERSITY

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

in

Environmental Systems

By

Otto Alvarez

Committee in charge:

Dr. Qinghua Guo, Chair
Dr. Thomas C. Harmon
Dr. Yihsu Chen
Dr. Robert C. Klinger

2015

The Dissertation of Otto Alvarez is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

Dr. Robert C. Klinger

_____

Dr. Yihsu Chen

_____

Dr. Thomas C. Harmon

_____

Dr. Qinghua Guo, Chair

University of California, Merced

2015

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would first like to thank my advisor Professor Qinghua Guo for his guidance and support in my research. He has been very supportive and has help me develop my skills to become a professional.

I would also like to thank Professor Yishu Chen, Professor Thomas C. Harmon, and Dr. Robert C. Klinger. They have all gave me constructive comments and guidance in my research during my stay at University of California, Merced.

I am also grateful to my colleagues: Paul Doherty, Jacob Flanagan, Wenkai Li, Mia Nelson, Yanjun Su, Andrew Zumkher, and all of the undergraduates who spend time working in the GIS lab. They also provided me with a lot of support in conducting my research.

Portion of Chapter 2 is a reprint of the material as it appears in (Alvarez, O., Q. Guo, R. C. Klinger, W. Li and P. Doherty (2013). "Comparison of elevation and remote sensing derived products as auxiliary data for climate surface interpolation." International Journal of Climatology). The co-author listed in this publication directed and supervised research which forms the basis for the dissertation. Permission to use copyrighted material in the dissertation has been granted by John Wiley and Sons.

Portion of Chapter 2 is currently under-review (Guo, Q. and O. Alvarez. A high spatial resolution climate surfaces utilizing remote sensing, to reduce uncertainty in interpolation and uncertainty surfaces. 2015). The co-author listed in this publication directed and supervised research which forms the basis for the dissertation.

Chapter 3 is currently under-review (Alvarez, O., Q. Guo, R.C Klinger., Y. Su, T. Harmon. Limits to spatial resolution of interpolated climate surfaces. 2014). The co-author listed in this publication directed and supervised research which forms the basis for the dissertation.

Chapter 4 is currently under-review (Alvarez, O., Q. Guo, R.C. Klinger, Y. Su, W. Li. Assessing the uncertainty in ecological niche modeling.2015). The co-author listed in this publication directed and supervised research which forms the basis for the dissertation.

Chapter 5 is currently under-review (Alvarez, O., Q. Guo, R. Sweitzer, M. Kelly. The impact of utilizing biophysical variables derived from Lidar to model the Pacific Fisher habitat in the Sierra National Forest. 2015). The co-author listed in this publication directed and supervised research which forms the basis for the dissertation.

Lastly but not least, I would like to thank my brother and grandmother for their full support and my mother Vilma Cuezzi she is the best mother anyone can ask for without her support I would not been able to accomplish all of my goals.

CURRICULUM VITAE

Otto Alvarez
School of Engineering
Environmental Systems Graduate Group
UC Merced, CA 95344
E-mail: oalvarez@ucmerced, oalvarez00@gmail.com

EDUCATION

University of California, Merced — Merced, CA
PhD, Environmental System (School of Engineering) — 2015
Dissertation: Understating climate and the impact on biodiversity
Advisor: Qinghua Guo

University of California, Merced — Merced, CA
M.S., Environmental System (School of Engineering) — 2011

University of California, Merced — Merced, CA
B.S., Computer Science and Engineering — 2008

RESEARCH INTERESTS

Spatial analysis, geospatial, gis, ecological niche modeling, climate change, and environmental modeling.

PEER-REVIEWED PUBLICATIONS (* Under Review, **under preparation)

Li, W., Q. Guo, C. Elkan, **O. Alvarez**. Modeling the probability of presence of species: an evaluation of the presence and background learning algorithm. *Global Ecology and Biogeography,* (2015*).

Alvarez, O., Q. Guo, R.C. Klinger, Y. Su, W. Li. Assessing the uncertainty in ecological niche modeling. (2015**).

Guo, Q. and **O. Alvarez**. A high spatial resolution climate surfaces utilizing remote sensing, to reduce uncertainty in interpolation and uncertainty surfaces. (2015**)

Alvarez, O., Q. Guo, R.C Klinger., Y. Su, T. Harmon. Limits to spatial resolution of interpolated climate surfaces. *International Journal of Climatology.* (2014*).

Xue, B., Q. Guo, **O. Alvarez**, J. Xiao, S. Tao, L. Li. Global patterns, trends, and drivers of water use efficiency from 2000 to 2011. *Ecosphere*. (2014, Accepted).

Alvarez, O., Q. Guo, R.C. Klinger, W. Li, P. Doherty (2014). Comparison of elevation and remote sensing derived products as auxiliary data for climate surface interpolation. *International Journal of Climatology* 34(7): 2258-2268. DOI: 10.1002/joc.3835

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 Modeling* 246: 60-67. DOI: 10.1016/j.ecolmodel.2012.07.025

Doherty, P., Q. Guo and **O. Alvarez** (2012). Expert versus Machine: A Comparison of Two Suitability Models for Emergency Helicopter Landing Areas in Yosemite National Park. *The Professional Geographer* 65(3): 466-481. DOI: 10.1080/00330124.2012.697857

Guo, Q., W. Li, H. Yu, **O. Alvarez**. Effects of Topographic Variability and Lidar Sampling Density on Several DEM Interpolation Methods. *Photogrammetric Engineering and Remote Sensing*. 2010. Vol. 76, No. 6, pp. 701–712. DOI: 10.14358/PERS.76.6.701

PRESENTATIONS/POSTERS

Alvarez, O., Guo, Q. High spatial and temporal resolution interpolated climate surfaces for California. The Wildlife Society's. Sacramento, CA. 2012

Doherty, P.J., Q. Guo, W. Li, **O. Alvarez**, J. Doke. Space-Time analyses for forecasting and understanding future incident occurrence: a case-study from Yosemite National Park Search and Rescue. GIScience. Columbus, Ohio. 2012

Klinger, R.C., **O. Alvarez**, B. Hatfield, M.L. Brooks, J.R. Matchett, C.E. Soulard. Spatial and temporal variability in alpine meadow condition and boundaries in the Sierra Nevada mountain range of California. The Wildlife Society's. Portland, OR. 2012

Alvarez, O., Q. Guo, R.C. Klinger. Alpine mammals and downscale method. NCCWSC Downscaling Workshop. Raley, NC. 2010

Whitenack, T., M. Williams, D. Tarboton, I. Zaslavsky, M. Durcik, R. Lucas, C. Dow, X. Meng, B. Bills, M. Leon, C. Yang, M. Arnold, A. Aufdenkampe, K. Schreuders, **O. Alvarez**. Building an integrated information system for publishing heterogeneous Critical Zone Observatory data. AGU. San Francisco, CA. 2010

Alvarez O., B. Doblack, and M. Dudys MOB2do. *Software Engineering (CSE 120)*,https://eng.ucmerced.edu/people/songhwai.oh/portal/teaching/cse120-sp08/CSE120_ClassPresentation.pdf 2008

GRANTS AND AWARDS

UC President's Dissertation Fellowship, UC Merced	2014 - 15
Bobcat Fellowship, UC Merced	2013 - 14
Environmental Systems Summer Graduate Fellowship, UC Merced	2010 - 14

TEACHING EXPERIENCE

Teaching Associate, Spatial Analysis and Modeling (45-60 students)	Fall 2008 - 11,13
Instructor, Spatial Analysis and Modeling (53 students)	Fall 2012

ABSTRACT

UNDERSTATING CLIMATE CHANGE AND THE IMPACT ON BIODIVERSITY
by
Otto Alvarez
Doctor of Philosophy in Environmental Systems
University of California, Merced, 2015
Dr. Qinghua Guo, Chair

Our understanding of how climate will change in the future is still very limited. There have been many studies conducted in trying to predict the future to obtain a rough idea of how climate will be. Understanding past climate is critical in trying to obtain an idea of how climate will be in the future. However, in order to analyze past climate, more accurate data needs to be generated. Climate surfaces, which are used in almost every environmental subfield, are critical to have. Most of the data available lack appropriate spatial attributes, such as, spatial extent and resolution, temporal resolution, and reduction in uncertainty. Therefore, one the objective of this research is to generate a dataset that will provide the necessary information to determine how climate has changed and the trends that have been experienced. A systematic evaluation was performed using different elevation and remote sensing products to improve the accuracy of climate surfaces. The results confirmed that remote sensing data significantly outperformed the commonly-used elevation product to generate climate surfaces, particularly for precipitation. This leads to determination of the optimal spatial resolution based on currently available weather data. Many high spatial resolution climate surfaces have been created without adequate understanding of how the generation of increasingly fine resolutions influences uncertainty. Findings show that regardless of the ecological zone, eco-region, or elevation zone, there were not any statistically significant differences among the uncertainties of all spatial resolutions. Although this indicates that interpolation of fine scale climate surfaces will generally not result in greater or lesser uncertainties, there will often be practical limits that dictate the logical limits of spatial resolution. For instance, the accuracy of a weather station location is often two or fewer decimal places (about 50% of the data), which makes the derivation of surfaces with resolutions < 1km inappropriate. With the appropriate climate surfaces a test was done to determine the uncertainty that has influence in ecological niche modeling. Ecological niche modeling is a popular tool that provides a probability of distribution for a species based on the connection made with the parameters that are fed in; yet, there has been a lack of consideration for uncertainty, especially for environmental surfaces. An experiment was conducted to determine the impact of uncertainty on species location and environmental surfaces. Utilizing the uncertainty information obtained from the climate

surfaces and the uncertainty of species location obtained from GBIF, we will be able to include uncertainty in the ecological niche models. The test species were 43 well-known distributions of mammals in the United States. By running Monte Carlo simulation and sensitivity analyses, the findings were that uncertainty in climate surfaces has a significant difference compare to the base than DEM. Uncertainty in the three aspects (topography, climate surfaces, and point locality) are critical information to be include and consider when generating a species model. Without this information one cannot conclude the probability of any single species.

CHAPTER 1: INTRODUCTION

Historic and localized weather data are used by climatologists to indicate climate patterns in the past and to make predictions for the future. Normally, the weather station data are sparsely distributed and could be considered as point data. In order to generate wall-to-wall gridded climate surfaces, an interpolation method is needed. Interpolation is a spatial analysis method using points in geographical (e.g. weather stations) and temporal space to predict climate variables in areas where there is no weather observation data (Daly et al. 2002, New et al. 2002, Hijmans et al. 2005, Mbogga et al. 2009). The products from these analyses are known as climate surfaces, and over the last decade they have been increasingly used in a range of studies, including ecology, hydrology, fire modeling, and water resources (Bonan et al. 2003, Kalnay and Cai 2003, Sheffield et al. 2004, Guo et al. 2005, Chen et al. 2007, Trabucco et al. 2008, Loarie et al. 2009, Mbogga et al. 2009, Thornton et al. 2009). Despite their wide use, there is a need for finer spatial and temporal resolution surfaces to make meaningful inferences at regional and monthly scales (Heikkinen et al. 2006). Temporal resolution is important for relating ecological responses to climate change, especially those associated with population dynamics (birth, death, and migration) of species whose lifespan and fecundity period varies from hours to decades (Walther et al. 2002). With regards to spatial resolution, recent climate surface development has focused on incorporating a high number of weather station data points, but has not fully accounted for the effect of spatial distribution on the validity and interpretation of their models (Thornton et al. 1997, New et al. 1999, Daly et al. 2000, Maurer et al. 2002, New et al. 2002, Hijmans et al. 2005, Allan and Ansell 2006).

A Digital Elevation Model (DEM) has been the covariate most frequently used in the development of climate surfaces. However, there has been little evaluation of additional covariates to complement DEM (Diodato 2005, Daly 2006). Due to the advances of remote sensing, GIS, and GPS technology, there is a wide range of spatially explicit data available as covariates that could be used to improve interpolation accuracy. Although previous research has suggested that adding such covariates could improve interpolation results, few studies have comprehensively examined the use of other elevation and remote sensing derived products for climate data interpolation (Hijmans et al. 2005). Moreover, lessons learned from geographic information science (GIScience) reiterate the importance of accounting for spatial uncertainty and unequal distribution during interpolation because some relationships between variables vary spatially (Fotheringham et al. 1998). In order to select the covariates, the product has to meet at least two basic criteria: (1) the data exist wall-to-wall in the study area and (2) fine spatial resolutions (e.g. $1km^2$) for most climate products. That is why remote sensing data are perfect to try to include as a covariate into creating climatic surfaces.

Even with creating the best climate surfaces and utilizing advanced remote sensing products, there are still limitations in creating such surfaces. There have been many climate products created in the last decade and they are all very unique (Daly et al. 2002, Hijmans et al. 2005, Hofstra et al. 2008, Alvarez et al. 2013). Currently, many of the products are being produced at a finer resolution without acknowledging or conducting an in depth analysis on the consequences of creating such a fine climate surface. This potentially can lead to unusable/unstable data, which creates a ripple effect mainly directed toward the management agencies and decision makers, which can make decisions or policy based on conclusion/finding on the product. Studying the limitation, if any, in creating such a fine product needs to be conducted at a global scale since the surfaces are widely used to make critical decisions such as in protecting species or their habitat.

The applications of utilizing climatic products are endless and are currently being used in many research fields. One field in which climatic products has been extensively used is in ecology, mainly in trying to understand the impact of climate change on a species (Feria and Peterson 2002, Guo et al. 2005, Li et al. 2011, Fernández et al. 2012). Niche modelling is widely used to determine the potential distribution of a given species. There have been many niche modelling algorithms developed or modified to be capable of handling such a task. In most of the cases when utilizing niche modelling climate, topography, and point locations are given to the model. However, one key problem or issue is that uncertainty is not considered in any of given data. This issue is critical since most model outputs are being interpreted without acknowledging this uncertainty. Trying to understand the impact of uncertainty in all the input variables in niche modelling is as important as trying to understand the potential habitat of the species; without fully understanding the impact of uncertainty in the input data, one cannot fully model the species realistically.

Creating a realistic model is always the main goal, however, typically many users of niche modelling only consider physical variables since those are available in all the parts of the world and because they are easily accessible (Peterson and Nakazawa 2008). However, there are other variables that could potentially be more influential or suitable because of the species habitat constraint. Biophysical variables are very hard to obtain, since each species is unique in terms of what type of biophysical parameters would be suitable. Many researchers believe that including those more meaningful variables could potentially created a more realistic model, since species do not only depend on physical parameters.

In summary, there are major uncertainty problems when creating a climatic product that have not fully been studied and many assumptions are being made based on these surfaces. Therefore, the objective of chapter 2 is to improve the current method utilizing all the weather station in the United States while reducing the uncertainty. Chapter 3, focus in understanding the limitation of the climate surfaces. Chapter 4 answers the important question in understanding the impact on niche modelling utilizing these climate surfaces,

and chapter 5, we determined the impact of including more ecological meaningful variables in niche modelling.

CHAPTER 2: GENERATING HIGH TEMPORAL AND LOW ERROR CLIMATE SURFACES FOR THE USA

INTRODUCTION

Climate surfaces are critical data needed to run any sort of models to predict either the future, past, or current climate. Many research fields use these surfaces as information to feed their model (Bonan et al. 2003, Kalnay and Cai 2003, Guo et al. 2005, Li et al. 2011, Fernández et al. 2012). Since weather stations are sparsely distributed, one cannot just simply use this information; most study locations are not around any stations (Alvarez et al. 2013). That is why climate surfaces are becoming very popular, to create such wall-to-wall gridded surfaces one needs to first obtain current/past weather station measurements (Daly et al. 2002, New et al. 2002, Hijmans et al. 2005, Alvarez et al. 2013). Then, an interpolation needs to be applied. Additionally, some interpolation algorithms allow the usage of covariates. Covariates are extra information to reduce the error/uncertainty; this has been adapted in almost all of the current climate surfaces available. They are widely used and accepted, however, the impotence of selecting the correct one is critical for each variable.

There are many climate surfaces available; be that as it may, there is a lack of providing a good model with the lowest uncertainty and the error/uncertainty associated with creating these products. WordClim (Hijmans et al. 2005) and PRISM (Daly et al. 2008) are two public climate surfaces available which are highly popular products. However, they do lack certain aspects such as: WorldClim's temporal resolution is an average from 1950 to 1999, PRISM spatial resolution is 4km (publicly available only), but most importantly, the lack of providing certainty information of the data. The uncertainty surfaces are as valuable of information as the interpolated data itself; without this critical information one cannot provide an accurate representation when utilizing this data. For example, in Ecological Niche Modeling (ENM) there have been huge improvements in reducing the error in almost all of the areas, except in the most important one, environmental layers.

In this chapter, we will be utilizing all the weather station available for the United States of America and surrounding areas. The climate surfaces will be created by using the best covariate and Thin Plate Spline as the interpolation algorithm. The method applied will be from Alvarez et al. (2013) which tests different covariates to determine the optimal covariate for each variable. For the creation of the uncertainty surfaces, the surface will be interpolated using a Thin Plate Spline (TPS) to keep the consistency with the climate surfaces. Temporal resolution will be from 1950 to 1999 at monthly scale, which includes four variables (total precipitation, maximum temperature, minimum temperature, and mean temperature). The spatial resolution will be of 1km x 1km. For each set of variables an

uncertainty surface will be provided, using a simple TPS and Co-Kringing interpolation algorithm.

POTENTIAL USES

As aforementioned, climate surfaces are being used in many areas/fields to better understand the outcome of the effects of climate change. The importance of developing climate surfaces while minimizing the error is critical, and the derived product of the uncertainty is as critical as the surfaces themselves. One area in which climate surfaces are popular is ecological niche modeling, trying to understand the effects of climate change on biodiversity. Understanding climate can lead to predicting where heat waves have and/or will happen, which is critical information for our society. In the summer of 2003, the increase in number of heat waves had a huge impact, such that the average mortality rate increased (Robine et al. 2008). Europe had more than 15,000 additional deaths caused by these heat waves, which makes it critical to understand the climate change/shift.

METHOD

The weather station information utilized to create the first version of ClimSurf at the United States extent are from Food and Agriculture Organization FAOclim2.0 (FAO 2001), the Global Historical Climate Network Dataset (GHCN) version 2 (Peterson and Vose 1997), and R-HydroNet version 1 (Vörösmarty et al. 1998). The number of weather stations are as follows: total precipitation 4 937, maximum temperature 3 053, minimum temperature 2 100, and mean temperature 3 646. All of these stations might not have been use for all the months since some months did not record the variable. The data were manually checked station by station. For temperature we graphed each station's yearly and monthly means for all of the years available and determined if there were any outliers by visual inspection. For precipitation, we detected outliers with a spatial outlier test, which included the precipitation information from the surrounding weather stations (within 250 km^2). Afterwards, we determined if the stations information was correct by comparing the latitude and longitude to its country/city and by comparing the given elevation value to that obtained from the USGS DEM layer. Most of the errors that were removed or corrected from stations were caused by incorrect units, equipment malfunction, and/or human error. Most of the errors were obvious and easily identified and corrected (e.g. where longitude and latitude for the station were swapped). These situations were generally apparent from the weather station data being markedly different from neighboring stations. Other weather stations had values that made sense for a period of time, but then appeared to have a

multiplier. This was due to the fact that the units for the weather station were changed. Most of the weather stations for our study area also reported elevation, so we were able to cross check the recorded elevation with our DEM layer to make sure that the weather station was not at the wrong location.

The selection of covariates and interpolation algorithm are based on the analysis done in Alvarez et al., (2013), which analyses multiple covariates, including remote sensing product, to determine which one was the best in reducing the uncertainty for each variable. Radar, rainfall data will be used as the covariate for precipitation since it had the lowest uncertainty for every month. Radar was available at the United States level by the National Oceanic and Atmospheric Administration (NOAA) National Weather Services (NOAA 2012). Since Radar is available at 4km, interpolation was applied to increase the spatial resolution to 1km utilizing Digital Elevation Model (DEM) (Farr et al. 2007) as a covariate (Alvarez et al. 2013). For maximum, minimum, and mean temperature, DEM was used as the covariate as it was found to be the best covariate. The interpolation algorithm, TPS (Furrer et al. 2011), was also applied to create the interpolation uncertainty surfaces. To generate climate surfaces, we used TPS in the package 'Fields' version 6.3 in R version 2.7.1 (Furrer et al. 2011). TPS aims to derive coherent signals and remove noise from an interpolation (Wahba and Wendelberger 1980, Wahba 1990), and was first applied in climatology by Hutchinson *et al*. (Hutchinson and Gessler 1994, Hutchinson 1995). The following equation is for TPS for two independent position covariates and extra covariates:

$$q_i = f(x_i, y_i) + \sum_{j=1}^{p} \beta_j \psi_{ij} + \varepsilon_i (i = 1, \dots, n), \quad (1)$$

and the smoothing function $f(x_i, y_i)$ and β_j are estimated by minimizing

$$\sum_{i=1}^{n} \left[\frac{q_i - f(x_i, y_i) - \sum_{j=1}^{p} \beta_j \psi_{ij}}{d_i} \right]^2 + \lambda J_m(f), \quad (2)$$

where $f(x_i, y_i)$ is the unknown smooth function, β_i is a set of unknown parameters, x_i, y_i, ψ_{ij} are the independent variables, ε_i is the independent random errors with zero mean and variance ($d_i \sigma^2$), $d_i$are the known weights, $J_m(f)$ is a measure of the smoothness of f defined in terms of mth order derivates of f, and λ is the smoothing parameter (Hutchinson and Gessler 1994, Hutchinson 1995).

To evaluate which covariate produced the lowest uncertainty for each climatic variable, a ten-fold cross-validation approach was used (Kohavi 1995, Hijmans et al. 2005). The climatic data are first divided randomly into 10 sub-samples and then TPS is run on 9 out of 10 sub-samples, retaining one sub-sample to validate the model. We repeated the process ten times, guaranteeing that all of the points are used for both training and validation. The model accuracy was then determined by the Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i^p - x_i^r\right)^2}{n}}, \quad (3)$$

where x_i^p and x_i^r are the model prediction and observed value for point *i*, and *n* is the total number of points, respectively. This is calculated monthly, having a total of 10 (runs cross-validation) by 12 (months) = 120 runs per climatic variable and covariate. After selecting the best covariate(s), all the points are then used to create the monthly average from 1950 to 1999 climate surfaces at a spatial resolution of $1km^2$. Then, using the same parameters, a monthly climate surface was generated for each year, creating 50 (years) x 12 (months) = 600 climate surfaces per climatic variable.

DATA FORMAT & UNITS

The climate and uncertainty surfaces are available at https://gis.ucmerced.edu/ClimSurf. There are multiple file formats to choose from: ESRI GRID, TIFF, and ASCII with a spatial resolution of 1km at monthly time periods for the average from 1950 to 1999 (Figure 1). However, upon request, data will be available for the early period (1930 to present), which includes both climate and uncertainty surfaces. The units are for total precipitation in millimeters and for temperature in Kelvin.

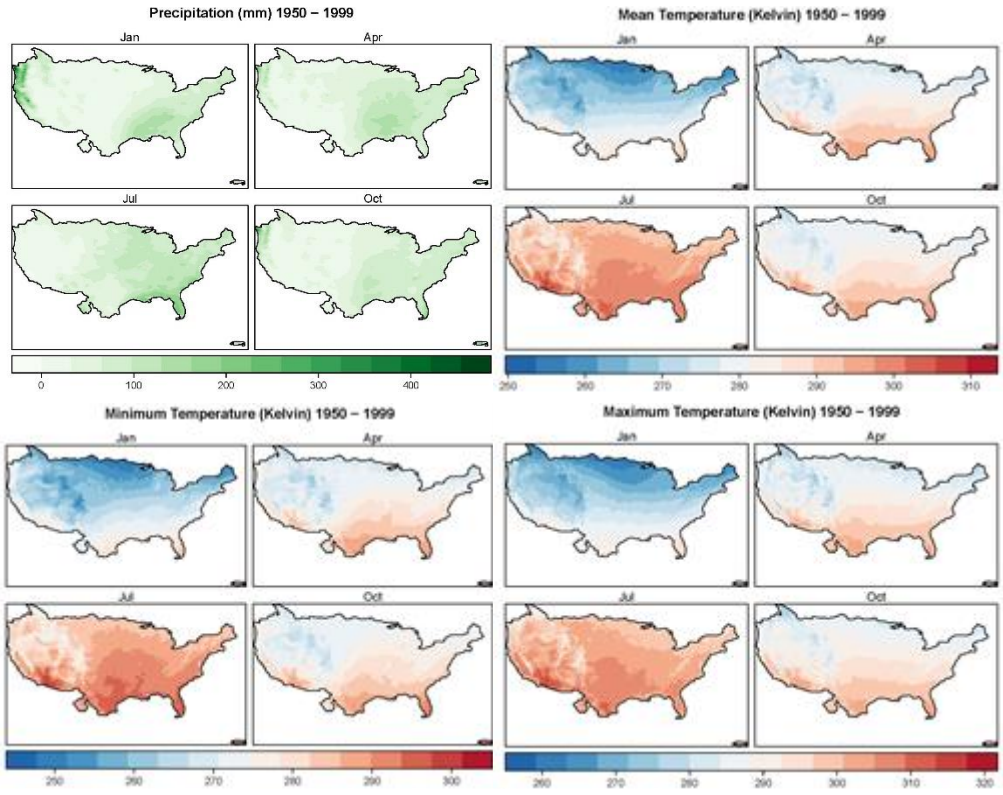


Figure 1: Summary climate layers from 1950 to 1999 for four months. Each set of quadrant represents a climate variable.

SUMMARY/DISCUSSION

It should also be noted that adding multiple covariates could increase rather than decrease the uncertainty in most cases, for all four climate variables. This can possibly be due to error propagation. Adding more covariates will add more sources of errors, which will propagate into the final product. Another possible reason is over-fitting. For the same sample size, adding more covariates can increase the risk of over-fitting. If we have more observation data, then the problem of over-fitting can be reduced.

Understanding past/present climate is critical in trying to obtain an idea of how climate will be in the future; nonetheless, in order to analyze past/present climate, better data needs to be generated. Climate surfaces, which are used in almost every environmental sub field, are critical to have. Key information which most of the climate products are not providing is the uncertainty information. Uncertainty surfaces are critical information that needs to be known before conducting any decision making, modeling, or anything that uses the climate surfaces. A set of climate surfaces and the uncertainty obtained from generation of those surfaces has been generated at a spatial resolution of 1 km.

CHAPTER 3: LIMITS TO SPATIAL RESOLUTION OF INTERPOLATED CLIMATE SURFACES

INTRODUCTION

Climatic surfaces are gridded data products that are generated by interpolating weather station information (Leemans and Cramer 1991, Daly et al. 2002, New et al. 2002, Hijmans et al. 2005, Hancock and Hutchinson 2006, Mbogga et al. 2009, Alvarez et al. 2013). These surfaces are used in many areas such as ecology, hydrology, fire modeling, species modeling, public health, and energy management (e.g., heat-wave implications)(Bonan et al. 2003, Kalnay and Cai 2003, Heikkinen et al. 2006, Chen et al. 2007, Robine et al. 2008, You et al. 2008, Loarie et al. 2009, Mandelik et al. 2010, Fernández et al. 2012). For many studies, a fine spatial resolution surface is critical for creating a good representative model, hence many regional climate surface products have been created with increasingly fine spatial resolutions (Daly et al. 2008, Haylock et al. 2008, Flint and Flint 2012). Many management agencies and decision makers believe that finer spatial resolution data will provide better information on how climate change will affect natural and anthropogenic systems (Brooks and Doswell 1993, Christensen et al. 2002, Damschen et al. 2010). However, critics believe that the creation of fine spatial resolution climate surfaces increases the error to a level at which the surfaces are unreliable (Christensen et al. 2002, Dessai et al. 2009). In this study, the error is defined as the difference of the interpolated data to the weather station data.

Among the first climate surfaces were those generated by New et al. (1999) and Leemans and Cramer (1991), with spatial resolutions of approximately 55 km, followed by a global surface created by New et al. (2002) with a resolution of 18 km. WorldClim, created by Hijmans et al. (2005), has a global spatial resolution of 1km; it had a high impact, subsequently becoming one of the most widely used and cited global products. In parallel with the development of these global surfaces, PRISM regional climate surfaces were created for North America by Daly et al. (2000), first with a resolution of 4km, and later with one of 800 m (Daly et al. (2008). Since then, many other regional scale climate surfaces have been generated; the assumption has been that finer resolution surfaces would lead to a better understanding of ecosystems and ecological patterns, resulting in better policy and management decisions (Wang et al. 2011, Flint and Flint 2012).

Despite these developments and the expected benefits of finer resolution surfaces, there has been insufficient research evaluating the degree to which increasing the spatial resolution leads to greater or lesser error in the surface. Most research has been done at the regional scale (Sharples et al. 2005, Flint and Flint 2012, Kearney et al. 2014, Sun et al. 2014), but good interpretation of regional scale climatic patterns requires an understanding

of how error varies across different regions at the global scale. However, we do understand that fine resolution of climate surfaces is needed to facilitate scale matching of other data (e.g. ecological data). Therefore, our primary goal was to determine whether increasingly fine scale interpolated climate surfaces also had increasingly great uncertainties. Relative to the spatial distribution of the weather stations we used to derive the climate surfaces, we were particularly interested in whether we would be able to identify a globally consistent spatial resolution with minimum prediction errors (uncertainties).

METHODS

Study area

The study was on 39 areas around the world (Figure 2). The areas were initially selected based on elevation, which we classified as low, medium, or high. We then selected areas for which the density of weather stations was > 40 per 300 km^2, which ensured that each location had an adequate sample for statistical testing. To ensure that the areas were representative of global ecoregions, we selected areas in all seven ecozones in the World Wildlife Fund's terrestrial ecoregion classification (Olson et al. 2001). This approach provided us with 35 areas for precipitation, 16 for maximum temperature, 19 for mean temperature, and 13 for minimum temperature.

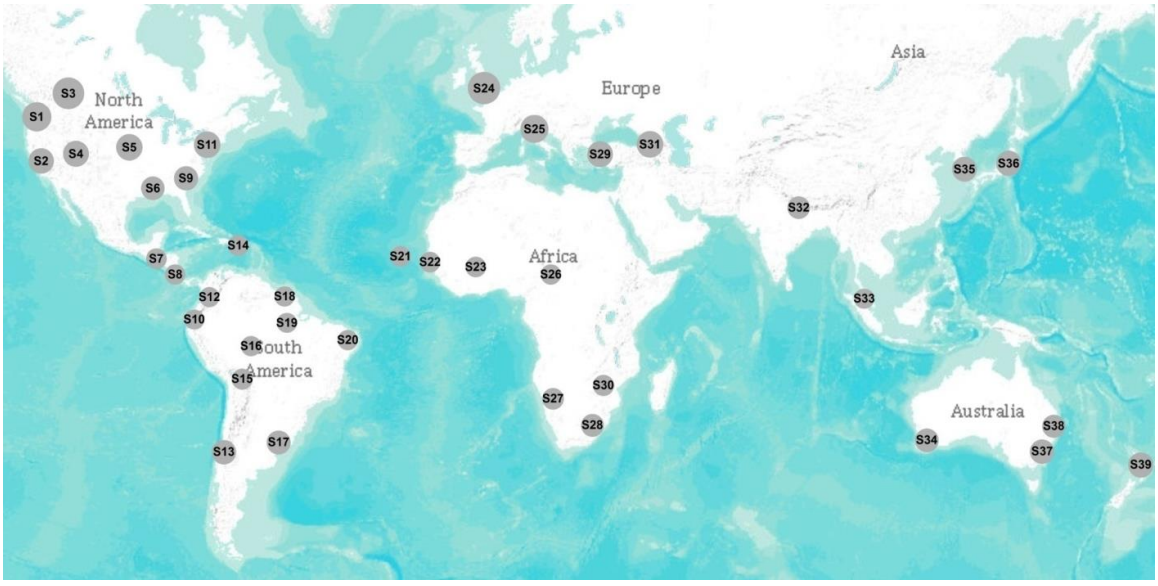


Figure 2: Study area consisted of 39 areas around the world, with 35 areas for precipitation, 16 for maximum temperature, 19 for mean temperature, and 13 for minimum temperature.

Weather station data and quality control

Weather station data were obtained from multiple sources/organizations, including the Food and Agriculture Organization FAOclim2.0 (FAO 2001), Electronic Hydrometeorological Data Network (R-HydroNet), and the National Climatic Data Center (gsod, ghcnv2). In total, there were 6 253 stations distributed among the 39 sites (3 370 for precipitation, 1 108 for mean temperature, 907 for maximum temperature, and 868 for minimum temperature), with a mean across all variables of ~75 stations per site (Table 1). The time period of 1950 to 1999 was selective because most of the popular surfaces and since this is the acceptable baseline. For each variable the average was taken for the entire period. The dataset was manually checked for any outliers (Alvarez et al., 2013). If multiple points are within a pixel, then we took the average of the values and assigned it to the center of the pixel to avoid the multiple point problem.

Table 1: Statistics (total count (#), average, minimum, maximum, and standard deviation) for each study area based on the weather station used. NA means that there is no data for that study area.

		Precipitation					Mean Temperature					Maximum Temperature					Minimum Temperature				
Area	ID	#	Avg	Min	Max	Stdv	#	Avg	Min	Max	Stdv	#	Avg	Min	Max	Stdv	#	Avg	Min	Max	Stdv
North America	NA1	94	28.9	5.9	71.8	16.5	65	32.7	5.9	77.9	18.5	52	37.9	5.9	106.3	22.2	58	35.4	5.9	106.3	22.1
	NA2	89	28.8	4.8	117.7	16.5	44	41.1	3.7	116.6	32.2	32	49.2	3.7	141.1	34.5	37	44.0	3.7	141.1	33.7
	NA3	44	45.2	4.7	96.8	25.1	32	48.0	4.0	130.7	31.3	31	50.5	3.1	119.2	35.7	46	41.3	6.0	120.4	22.1
	NA4	36	39.7	5.7	119.6	23.2	45	43.6	6.0	120.4	24.2	43	43.9	6.0	120.4	24.9	68	32.9	1.9	78.2	20.3
	NA5	56	37.0	1.3	73.9	17.4	69	31.7	1.9	73.9	20.3	58	37.2	1.9	112.4	23.9	NA	NA	NA	NA	NA
	NA6	72	35.2	5.4	106.5	18.9	59	41.3	3.8	107.1	20.0	30	54.7	3.8	141.5	33.9	57	41.5	3.8	107.5	20.1
	NA7	84	31.6	4.5	83.5	18.3	71	36.3	4.5	83.2	19.4	66	41.6	4.5	83.2	17.6	66	41.6	4.5	83.2	17.6
	NA8	125	24.8	3.7	57.5	13.3	101	27.7	4.0	105.3	16.9	94	29.0	4.1	104.9	17.0	99	28.2	4.1	104.9	16.7
South America	SA1	167	14.6	1.9	82.8	12.0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	SA2	145	12.5	1.9	94.3	10.5	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	SA3	112	18.5	1.9	62.4	13.9	29	30.5	3.1	124.0	28.3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	SA4	75	18.8	4.0	84.5	13.1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	SA5	49	31.4	6.7	213.8	40.2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	SA6	74	20.2	2.6	137.7	25.9	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	SA7	68	24.8	2.2	128.2	23.6	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	SA8	49	39.3	5.6	80.2	22.4	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	SA9	NA	NA	NA	NA	NA	22	55.7	3.1	174.0	49.3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	SA10	45	25.7	5.6	174.1	34.7	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	SA11	52	43.6	4.0	101.7	23.3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	SA12	318	15.2	1.9	46.7	9.8	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Europe and China	EC1	51	37.8	6.7	102.2	21.2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	EC2	NA	NA	NA	NA	NA	36	32.2	2.6	93.1	27.8	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	EC3	54	43.9	13.4	91.1	15.7	87	23.0	1.9	97.6	21.2	84	22.3	1.9	97.6	21.3	84	22.3	1.9	97.6	21.3
	EC4	44	49.6	8.3	88.0	19.3	38	36.7	2.6	159.9	37.2	36	39.1	2.6	159.9	38.5	36	39.1	2.6	159.9	38.5
	EC5	67	31.3	5.6	127.9	22.3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	EC6	62	31.3	10.5	104.5	15.9	62	32.6	5.8	104.6	15.6	36	43.5	11.5	104.5	20.1	36	40.7	11.0	80.0	14.7
	EC7	NA	NA	NA	NA	NA	37	47.3	3.3	95.4	25.5	41	47.6	3.3	89.7	21.3	NA	NA	NA	NA	NA
Africa and Australia	AA1	57	10.8	3.7	41.5	7.6	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	AA2	122	20.6	1.9	53.5	11.3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	AA3	142	28.4	2.6	101.3	12.5	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	AA4	88	24.5	4.1	74.9	17.2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	AA5	59	39.1	1.9	127.6	23.0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	AA6	111	25.3	2.6	91.6	19.2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	AA7	129	23.3	1.9	74.6	12.0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	AA8	43	34.9	13.4	95.2	18.2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	AA9	NA	NA	NA	NA	NA	28	57.5	6.5	181.0	36.5	27	60.1	22.9	180.7	34.7	NA	NA	NA	NA	NA
	AA10	71	32.9	2.6	93.2	21.4	55	38.5	6.2	93.6	24.8	53	42.9	7.9	104.4	26.1	53	42.9	7.9	104.4	26.1
	AA11	51	36.9	8.3	91.3	20.0	49	34.6	1.9	114.0	23.6	45	33.7	1.9	114.0	24.0	46	33.9	1.9	114.0	23.8
	AA12	465	8.8	0.8	72.9	5.2	179	14.4	0.5	52.5	9.4	179	14.6	0.5	52.5	9.4	182	14.5	0.5	52.5	9.3

The climate surfaces were evaluated using a traditional ten-fold cross-validation (Kohavi 1995, Hijmans et al. 2005, Alvarez et al. 2013). The approach was as follows: weather station data for each variable and area were randomly divided into 10 subsamples and then run through the model for nine of the 10 samples. This guarantees that each point will be used at least once for training and validation of the model. We repeated this process 10 times and used the root-mean-square error (RMSE) to calculate the accuracy.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i^p - x_i^r\right)^2}{n}}, \tag{4}$$

where x_i^p and x_i^r are the model prediction and observed values for point *i*, and *n* is the total number of points, respectively.

The interpolation algorithm we used to create the climate surfaces was Thin Plate Spline (TPS). TPS is widely used in creating climate surfaces because it can derive coherent signals and remove noise when determining values for each cell (Wahba and Wendelberger 1980, Wahba 1990, Hutchinson and Gessler 1994, Alvarez et al. 2013). The algorithm we used was from the "Fields" package, version 6.8 in R version 3.0.1(Furrer et al. 2011). The TPS equation for two independent position covariates and extra covariates is

$$q_i = f(x_i, y_i) + \sum_{j=1}^{p} \beta_j \psi_{ij} + \varepsilon_i (i = 1, \dots, n). \quad (5)$$

The smoothing function $f(x_i, y_i)$ and β_j are estimated by minimizing

$$\sum_{i=1}^{n} \left[\frac{(q_i - f\left(x_i, y_i - \sum_{j=1}^{p} \beta_j \psi_{ij}\right)}{d_i} \right]^2 + \lambda J_m(f), \quad (6)$$

where $f(x_i, y_i)$ is the unknown smooth function, β_i is a set of unknown parameters, x_i, y_i, ψ_{ij} are independent variables, ε_i is the independent random error with zero mean and variance ($d_i \sigma^2$), d_i is the known weight, $J_m(f)$ is a measure of the smoothness of *f*, defined in terms of the *m*th order derivatives of *f*, and λ is the smoothing parameter (Hutchinson and Gessler 1994, Hutchinson 1995).

The covariate we used in the TPS (or co-kriging) was a digital elevation model (DEM) from the Shuttle Radar Topography Mission (STRM) at a spatial resolution of 90 m (Consultative Group on International Agricultural Research Consortium for Spatial Information) (Jarvis et al. 2008). The DEM was aggregated using the Raster package in R with mean parameters to obtain resolutions of 500m, 800m, and 1km. The aggregating needed to be done since the spatial resolution of DEM has to match our target climate surfaces. DEMs are used in many climate surfaces such as WorldClim, PRISM, and ClimSurf (Hijmans et al. 2005, Daly et al. 2008, Alvarez et al. 2013).

Tukey's statistical test, which is a robust single-step comparison procedure for finding means that are significantly different from each other, was used to determine whether there were significant differences among the RMSEs of the four resolutions (Duncan 1955).

There are three assumptions in Tukey's test: 1) the observations being tested should be independent within and among groups; 2) the within-group variances across the groups associated with each mean should be equal; and 3) each observation being tested should be normally distributed. Tukey's test is essentially a t-test, but it can reduce the experiment-wise error rate (type I error) (Duncan 1955). A null hypothesis for Tukey's test shows that all the means being compared are identical to each other. If the probability (*P*) of obtaining a null hypothesis is smaller than the prescribed significance level (α) of 0.05, this null hypothesis will be rejected, indicating that there are significant differences among the means. Otherwise, the null hypothesis is accepted. The null hypothesis was that the mean RMSEs of the four resolutions were similar to each other. The *P*-value for the test was calculated from the studentized range, q_s,

$$q_s = \frac{|\mu_a - \mu_b|}{s}, \tag{7}$$

where μ_a and μ_b are the means from two different groups being compared, and s is the standard deviation of the sample. The assessment was performed in multiple ways. The test was performed by region for each month, then all the stations for each month, and then repeated in the same manner, but annually, for each variable.

RESULTS

Spatial resolution and inter-regional comparisons

The RMSE values across all 39 sites for precipitation, mean temperature, and minimum temperature were very similar for all spatial resolution classes (Figure 3). However, specific RMSE patterns were observed for some sites. For instance, precipitation at site NA8 had an error of 22mm at a resolution of 1km; the error then slowly decreased with increasing resolution, suggesting that, in some cases, the error is lower for a finer resolution. NA8 is located on the east coast of the United States (Pennsylvania, New York, and Delaware) and was classified as a low-elevation zone. The error in the NA3 maximum temperature at a resolution of 90m was three times greater than those for the other resolutions in North America, Europe, and China (Figure 3). The opposite is observed for Africa and Australia (AA8 and AA10-12), where the error in the maximum temperature at a resolution of 90m was as much as three times less than those of the other resolutions. Elevation classification did not affect the RMSE at any resolution, the 'constant' pattern remains. The Tukey's test results suggest all the differences between the mean of each

resolution group is statistically insignificant ($P > 0.99$ at $a = 0.05$); the results were consistent for all comparisons.

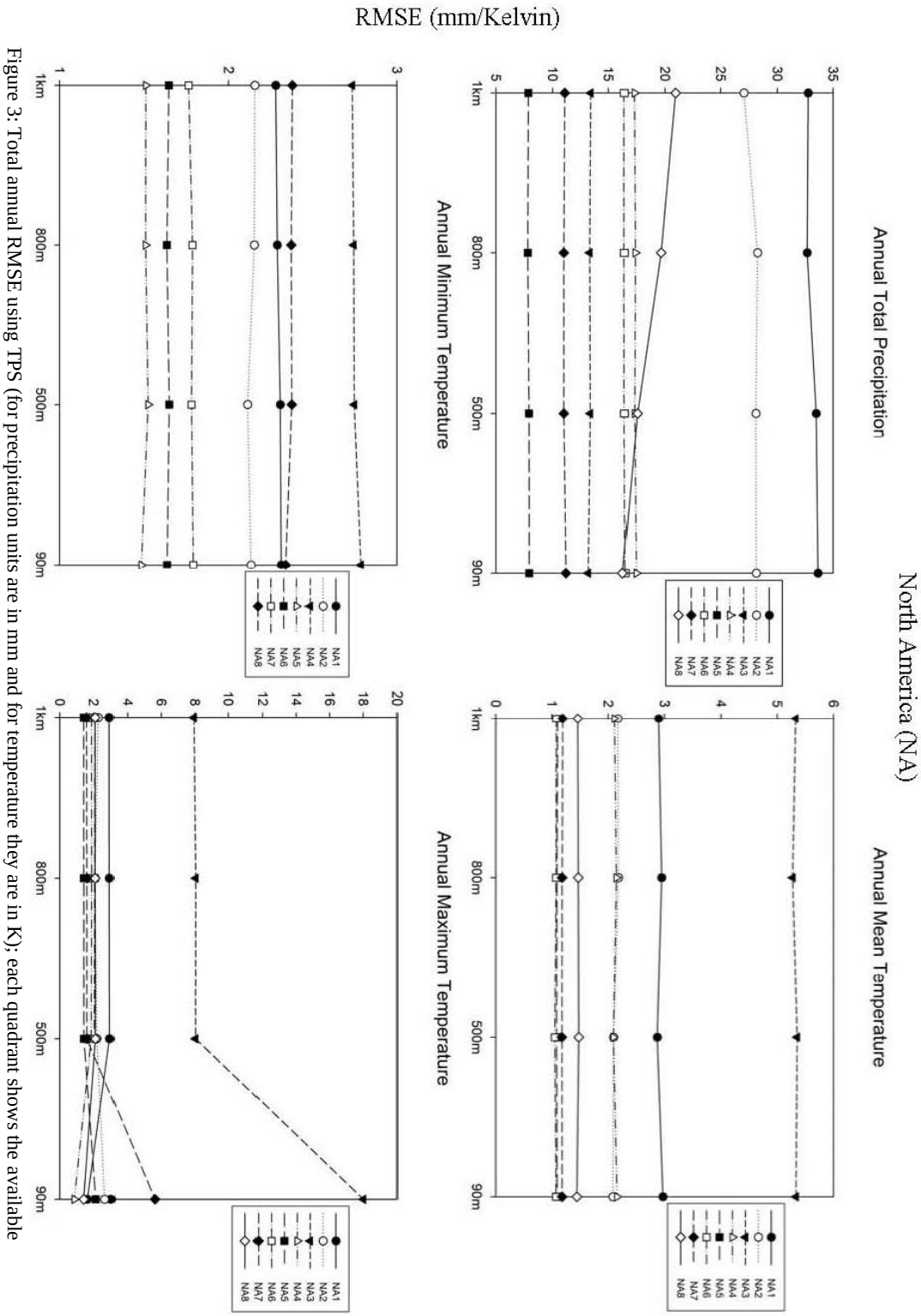


Figure 3: Total annual RMSE using TPS (for precipitation units are in mm and for temperature they are in K); each quadrant shows the available variables and display search study area in finer detail.

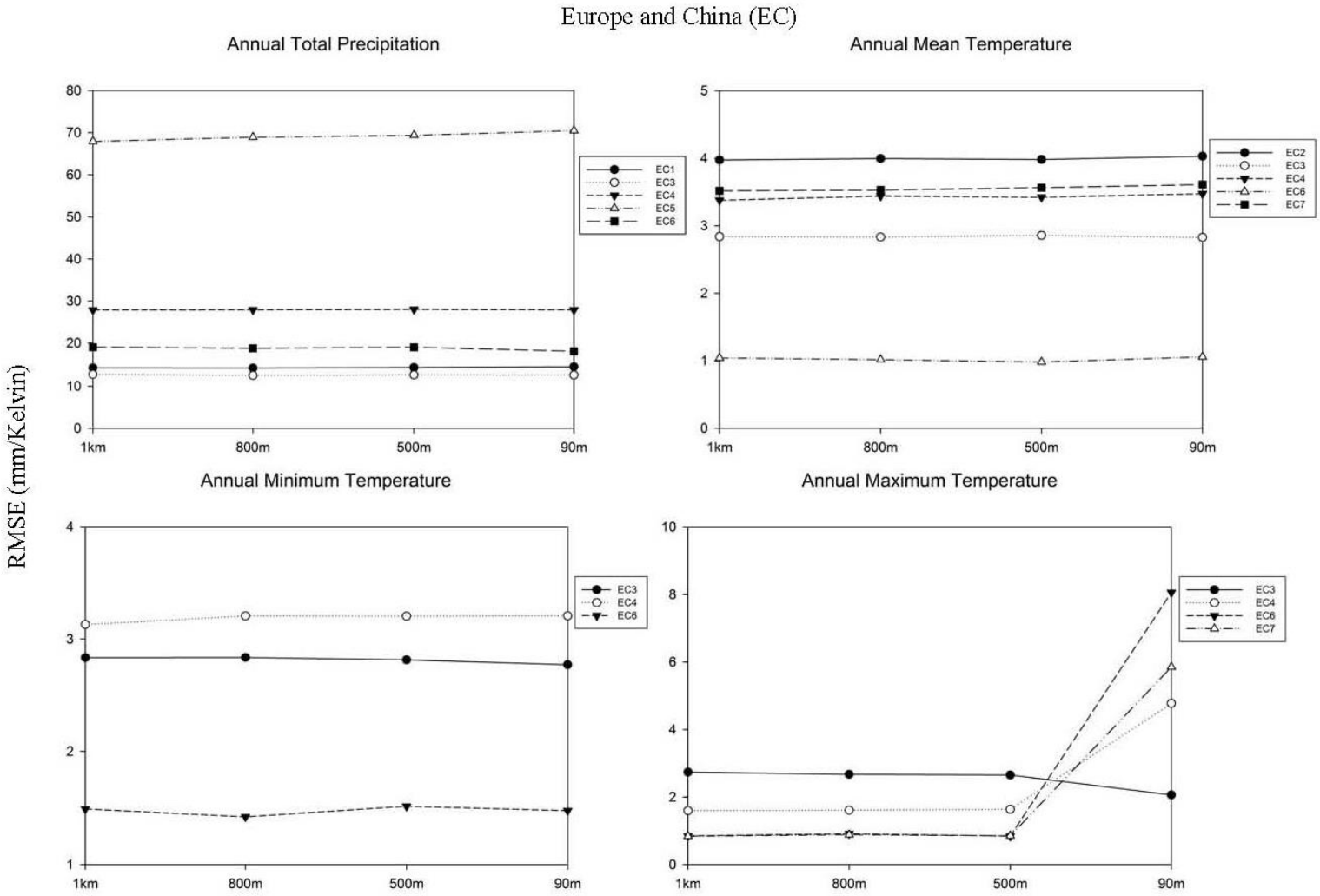


Figure 3: (Cont. Quadrant II):

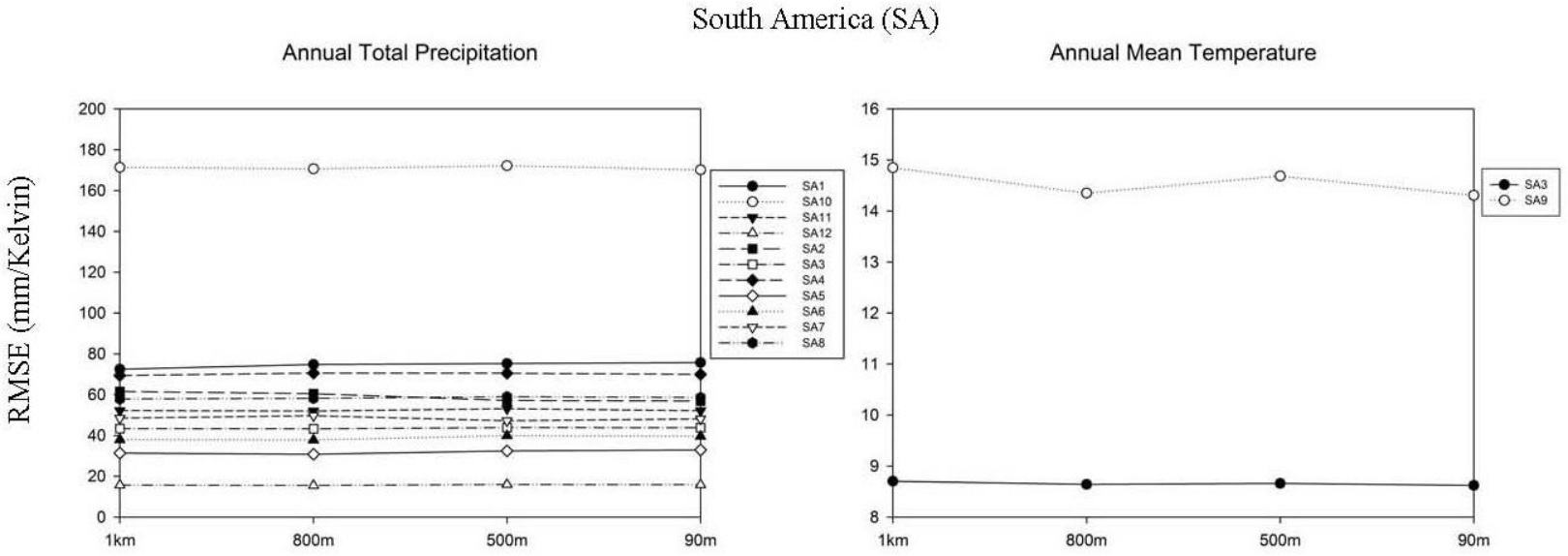


Figure 3: (Cont. Quadrant III):

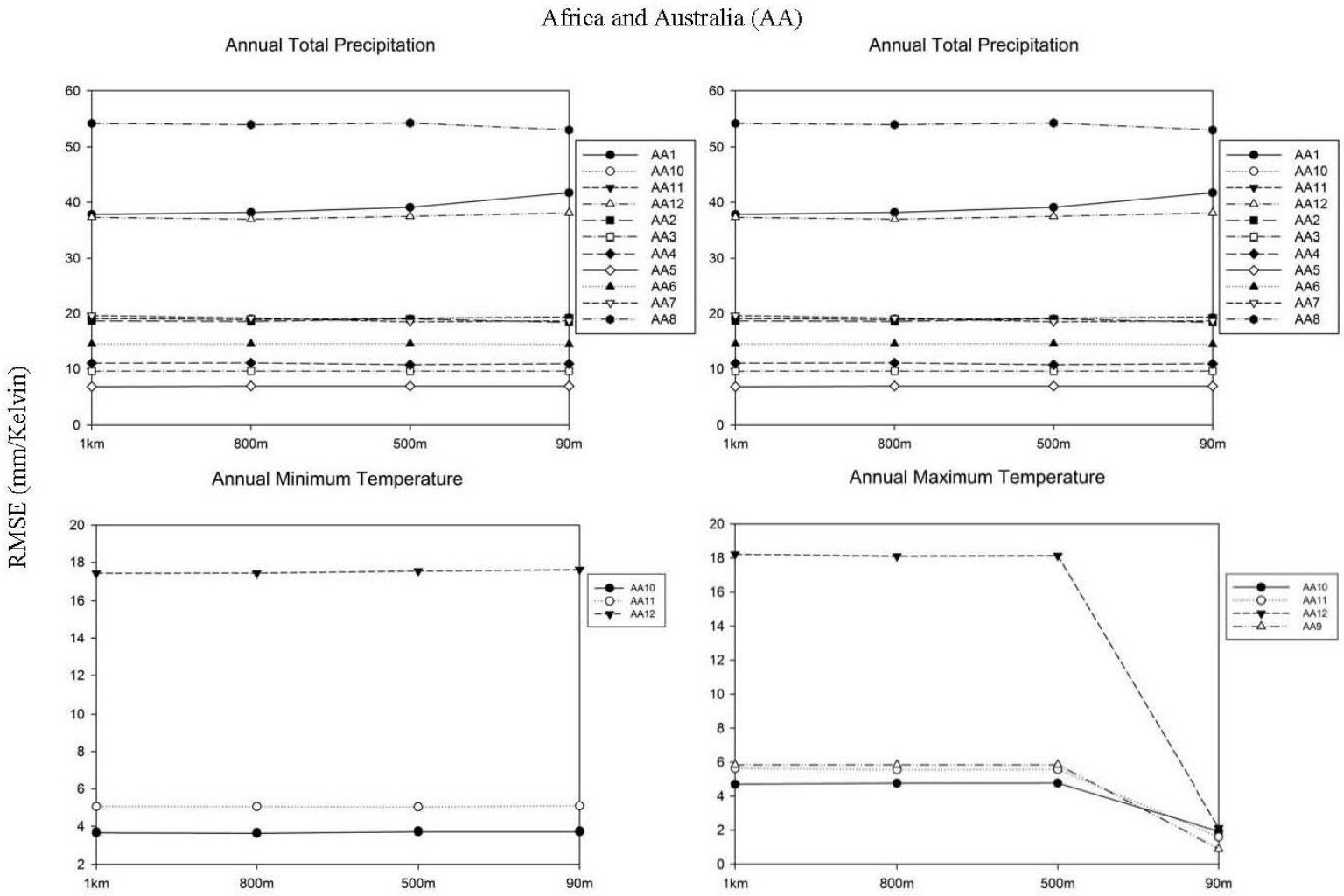


Figure 3: (Cont. Quadrant IV):

Monthly comparison by regions

Precipitation

In North America, the RMSE ranges were about two-fold larger for January, February, March, July, November, and December (Appendix A Figure 1) than for the other months. In general, the months that were highly clustered, meaning low variations, in North America were June, July, August, and October. The median for nearly all the months in North America were in the same range. In Europe and China, the RMSE ranges for June, July, August, and September were three times greater than those for the other months, but, overall, the RMSE was highly variable; there was no clustering for any of the months; the magnitude was higher in some months, but the variations were still significant. Again, as in North America, the median value for all the months was "consistent". In South America, the RMSE ranges were much greater than those in other regions, as much as four times greater for January, May, June, July, and August. There were large variations for most months, and the median was not "constant" as in the case of North America. In Africa and Australia, the RMSE ranges were in general stable, with the exception of two months. Again, almost all the months were clustered together, and for May, June, July, September, October, November, and December, we observed high extreme values compared with the median.

Mean temperature

In North America, the RMSE ranges for almost all months were very similar, with the exception of January and December (Appendix A Figure 2).The maximum values for January and December in North America were almost four times higher than those for the other months; however, most of the months were clustered together, with the exception of December, and the median values were similar. In Europe and China, the range of RMSEs was not large, meaning that there was not much difference, as in the case of North America. The variations were almost the same, but the median did change, depending on the month; for January to August, the median was twice that for September to December. We were unable to draw any conclusions for South America, because of the limited number of study areas in that region (Table 1). In Africa and Australia, the pattern was the same for each month: the range and median were constant and the variations high

Maximum temperature

The ranges for North America were very similar, except for January and February, for which the values were four times lower than those for the other months (Appendix A Figure

3). The months from March to December had very high, almost identical ranges of around 12. Half of the months, January to June, were clustered together, but variations were seen for July to December. The medians for the colder months, i.e., January, February, March, April, October, November, and December, were almost identical, around 1 K, and for May to September the median was around 2 K. In Europe and China, the range for August to December was twice that for January to July. For almost all months, there were large variations, but the medians were similar. In Africa and Australia, the RMSE range for almost every month was essentially the same, with a differences of a few degrees. The only exception was March, which had a lower range, and the maximum value was about 4 K lower than the average. All the months had high variations and the median was the same for almost all the months, except November and December, for which the median values were twice as high.

Minimum temperature

In North America, the RMSE ranges were constant for all months, but for January, February, March, June, July August, September, and December, the variations were high (Appendix A Figure 4). The median was also the same across all months, with the exception of February, for which the value was half those of the others. In Europe and China, the RMSE range for January to March was much lower than the range for April to December; also, the variations for the months with lower ranges were almost 2 K smaller. The median values and resolutions were different for each month .For Africa and Australia, the RMSE ranges were very high, with the exception of September, for which they were two-fold smaller. All the months had high variation and the median values were almost the same, with the exceptions of September and October.

DISCUSSION

Impact of resolution

Our analysis indicated that spatial resolution did not have a significant influence on the uncertainties in interpolated estimates of precipitation or temperature variables. This pattern held regardless of the region or month; there were idiosyncratic differences among the RMSE values, but these were not systematic. These findings are important because there have been strongly contrasting opinions on whether increasing the spatial resolution would provide better quality data for making decisions or whether finer resolutions would create unusable data. However, neither of these viewpoints is supported by our study. We found that the error when interpolating from finer resolutions does differ from those for coarser ones, and this did not vary substantially, irrespective of geographic region,

elevation zone, or ecoregion. The same analysis was conducted using co-kriging as the interpolation algorithm to test if our results would be sensitive to the choice of interpolation methods. Co-kriging was ran with the same covariable (DEM) based on the "Fields" package in R version 3.0.1(Furrer et al. 2011).The results showed little difference from TPS for annual, monthly, and regional analysis (see an example in Appendix A Figure 5). Based on our results, the similarities of co-kriging and TPS suggest that interpolation algorithm does not have much influence, when utilizing the same covariate.

These findings also have practical implications for how climate surfaces are developed. The process of creating increasingly fine resolution surfaces is computationally expensive, but instead of using interpolation to develop the surfaces, an alternative approach could simply be regridding of the data to the target resolution. However, even if this approach proves to be useful, weather station locations have inherent limitations that put constraints on the finest scale for which it is logical to develop surfaces. The minimum number of decimal places of the longitude and latitude of stations is frequently two digits, which results in an uncertainty of approximately 1 km (Wieczorek et al. 2004). For example, consider creating 100m climate surfaces with weather station locations with an uncertainty of 1km;the location of the station could be anywhere in a 1000m by 1000m region. In our dataset, about 55% of the weather station locations had two decimal places precision or less. It would therefore make little sense in these instances to attempt to create surfaces with resolutions much less than 1 km.

Terrain and other classifications

Other than weather station density, we had no prior expectations of how geographic location would affect our analyses, so we selected our locations based on data availability. We found that elevation zone, ecozone, and ecoregion had no effect on the error of the spatial resolution (Figure 3), but there were some exceptions to the general patterns, suggesting that further investigations would be useful. For example, when considering a single study area based only on changes in resolution, virtually every station had minimal variations, therefore the interpolated values were almost constant. An exception was precipitation in area NA8, which is located on the east coast of North America. It had the highest number of weather stations in that region as well as the lowest average, minimum, maximum, and standard deviation of all the study areas in North America. Moreover, it was classified as low elevation, which implies that elevation variations is minimal. Another example was the maximum temperatures in North America, Europe, and China, error was greatest at the finest resolution for most study areas in each region. However, in Africa and Australia, we observed the opposite pattern, with the error being lowest at the finest resolution. A possible explanation for this difference is that there is a high number of stations (179) in Africa and Australia while in North America, China, and Europe there are

few stations (40-60). This important difference can indicate that if the study area is rich in the number of stations, more than 100, than creating a finer climate surface will actually be beneficial. Theoretically it makes sense, the more information the algorithm is given the better the product and especially if the target surfaces will be produced at a fine spatial resolution (< 1km). The problem is that most of the stations are not near each other, therefore, this causes the problem of creating a product with a fine resolution (< 1km). However, results suggest that for precipitation this does not apply, since most of the study areas have high number of stations (100-160) while remaining constant in all resolution types, but it might be a variable independent issue. Again, based on the terrain classification, there was no consistent pattern, but these exceptions suggest that other factors may influence the error in as yet undetected ways.

Improvements and conclusion

There are several questions that should be addressed to improve our understanding of how error is affected by the creation of increasingly fine climate surfaces. As mentioned above, one particularly important one is how precision in the location of the weather stations influences the feasibility of generating climate surfaces < 1km in resolution. Another consideration is selection of covariates for generating the surfaces. We used DEM as a covariate when we generated our climate surfaces so we could make direct comparisons among regions, but in previous research we found that DEM was not the best covariate to use when generating precipitation surfaces (Alvarez et al. 2013). Lastly, we were not able to analyze certain regions since data were not available. For example, in South America we lacked maximum and minimum temperature, and in places like in China we lacked density of weather stations was less than desired for all variables. DEM was use in this study as a covariate since we wanted to keep as many parameters as possible constant in our analysis; the use of a more appropriate covariate could affect the global comparison.

The uncertainty of DEM is not considered in this study. As mentioned previously, DEM was used as the covariate for generating the climate surfaces. However, there are uncertainties in DEM, which vary significantly from bare earth to vegetated mountainous regions (Rodriguez et al. (2006); Su and Guo (2014). DEM uncertainty is a critical aspect to understand since it has a big influence the climate surfaces. Su and Guo (2014) found that the USGS DEM was systematically overestimated in vegetated mountain areas when compared to the Lidar derived DEM. In future studies, we plan to study the influence of DEM uncertainties on modeling climate surfaces.

In conclusion, we found that generating increasingly fine solution up to 90 meters climate surfaces neither increases nor decreases error. The accuracy of the location data is critical in determining whether increasing the target resolution will give a better surface. We

showed that there were no statistically significant differences among the uncertainties for meaningful ecological and geographic classes, so it may be unnecessary to generate finer solution climate surfaces, given the current quality of data on weather station locations.

CHAPTER 4: ASSESSING THE UNCERTAINTY IN ECOLOGICAL NICHE MODELING: LOCALITY, CLIMATE, AND TOPOGRAPHY

INTRODUCTION

Ecological niche models (ENM), has been a key in understanding the impact of climate change (Feria and Peterson 2002, Graham et al. 2004, Guo et al. 2005, Rodriguez et al. 2007, Peterson and Nakazawa 2008, Lenoir et al. 2010, Fernández et al. 2012). ENM builds a statistical relationship between the occurrences of the species to environment predictors, allowing it to predict future distribution under climate change and topography layers. There are many niche modeling algorithms that have been implemented: BioClim, Genetic Algorithm for Rule-set Production (GARP), Support Vector Machine (SVM), Presence and Background Learning Algorithm (PBL), Maximum Entropy (MaxEnt), and Generalized Linear Model (GLM) (Busby 1986, Carpenter et al. 1993, Stockwell and Peters 1999, Guisan et al. 2002, Guo et al. 2005, Phillips et al. 2006, Li et al. 2011). ENMs are not limited to predicting species distribution; they have also been used to predict other scenarios, such as classifying land cover to remote sensing images and determining helicopter suitable landing areas (Li and Guo 2010, Li et al. 2011, Doherty et al. 2012). Unfortunately, there are many issues related with uncertainties in the data inputted to generate the predictions. A major step forward has been achieved by removing the bias/uncertainty in training the model by only selecting background points from sampling areas (Elith et al. 2006, Phillips et al. 2009).

To create ENM, certain data need to be given to a model regardless of which algorithm is used. First, one needs to give the location of either of the species, present only or both present and absence. Second, environmental layers, which are typically climate surfaces and topography (i.e. elevation, slope, mean, maximum, minimum temperature and precipitation) (Peterson and Nakazawa 2008). Climate surfaces are generated using weather station data and applying an interpolation algorithm, likely using Thin Plate Spline (TPS) or Kriging (Hijmans et al. 2005, Hong et al. 2005, Hancock and Hutchinson 2006, Stahl et al. 2006, Hofstra et al. 2008, Alvarez et al. 2013). Topographic layers/surfaces in large spatial scale are generally calculated or derived using airborne photogrammetric technique (Honkavaara et al. 2009), radar technique (Farr et al. 2007) or Light detection and ranging (Lidar) technique (Guo et al. 2010).

Although there has been a lot of effort in trying to reduce the uncertainty in creating such data (point and surfaces), the error still exists. Locality data depends on the GPS accuracy at the time the samples were taken and on how many satellites are connected to obtain those coordinates. Also, if the data was geo-referenced, this will have an associated error/uncertainty. This area, however, has been already established (Guo et al. 2008) and

many ENMs studies utilizing this have been published (Fernandez et al. 2009, Newbold 2010, Pyke and Ehrlich 2010). With the development of GPS devices and differential GPS algorithms, the uncertainty of geo-locations for in-situ measurements has been dramatically decreased (Wing et al. 2005, Misra and Enge 2006). However, the positional accuracies for measurements taken in forest, which is the most frequent case for ENM, can be 100% percent lower than in open sky conditions (Wing et al. 2005). The uncertainty of frequently used topography surfaces has also been well defined. For example, the Shuttle Radar Topographic Mission (SRTM) digital elevation model (DEM) product, as one of the most frequently used near-global scale topographic products, has been carefully evaluated either regionally or globally (Bourgine and Baghdadi 2005, Berthier et al. 2006, Rodriguez et al. 2006, Weydahl et al. 2007, Su and Guo 2014). The vertical error for SRTM data is smaller than 1 meter for non-vegetated low-relief conditions (Rodriguez et al. 2006). However, with the increase of vegetation cover and relief of terrain, the vertical error of SRTM data can be over ten times higher (Su and Guo 2014). Climate surface uncertainty/error on the other hand, has not been a factor in ENM and there are very limited studies that have included such a factor. Climate surfaces is still a work in progress and since some of the variables are unpredictable, extra information is need. Climate surfaces are generated by utilizing weather station and interpolating, where covariates are usually added to the interpolation algorithm to reduce the uncertainty (Hijmans et al. 2005, Daly 2006, Alvarez et al. 2013).

There are very limited studies that include all of these (point, topography, and climate) uncertainties in trying to predict species niches. This is unfortunate because by understanding all of the uncertainties that propagate from the algorithms and the surfaces, there is a higher probability that we can build models that can better represent the niche or distribution of species. Most efforts are going into understanding the future niche uncertainty (Heikkinen et al. 2006, Morin and Thuiller 2009). There was, however, a study done by Fernández et al. (2013) which partially included uncertainty on current climate. The climate uncertainty was not produced during the time of the surface creation, therefore, the uncertainty produced by Fernández et al. (2013) may have some bias or "error" associated with it. Therefore, our goal in this study was to use a series of Monte Carlo simulations to evaluate the relative influence of uncertainty in species locations, contemporary climate, and topography on the predicted distributions for 43 North American mammal species.

METHOD

Study area and point localities

We extracted point localities of 43 non-volant mammal species from the Global Biodiversity Information Facility (GBIF). Not only does the GBIF contain millions of location records from a range of taxa on around the world, but it also has uncertainty information for each location (GBIF 2014). The methodology used to develop spatial data on climate and topography can vary among countries and regions, resulting in Geographic Information System (GIS) layers that vary in resolution, extent, and accuracy. Therefore, to ensure consistency we restricted our analysis on species locations to within the coterminous 48 states in the United States of America; these species were selected because it meets the following criteria: 1) at least 100 unique locations which were 1km apart, 2) the uncertainty was recorded for each point and it had to be bigger than 5km, and 3) species needed to be well established in the United States. Each species was then classified based on their potential habitat niche: board, moderate, and narrow.

Climate surfaces & topography

For the climate surfaces we used ClimSurf (Alvarez et al. 2013), created using observational weather stations from Food and Agriculture Organization (FAO) and the Global Historical Climate Network (GHCN) version 2 (Peterson and Vose 1997, FAO 2001) using TPS, which is highly used in creating climate surfaces (Hutchinson and Gessler 1994, Hutchinson 1995, Price et al. 2000, Jeffrey et al. 2001, Hijmans et al. 2005, Hong et al. 2005, Hancock and Hutchinson 2006, Hancock and Hutchinson 2006, Tait et al. 2006). ClimSurf also provides uncertainty information, which can be used to generate layers with uncertainty. The uncertainty information comes from the ten-fold cross-validation done in Alvarez et al. (2013).

The SRTM global DEM product was used to represent topographic surface in this study. SRTM is a joint mission conducted by NASA (National Aeronautics and Space Administration) and NGA, which was flown in February 2000. The C-band sensor on board the SRTM, which is used to generate global DEM product, covers approximately 99.97% of the Earth land surface from 56°S to 60°N at least once during an 11-day mission (Farr et al. 2007).The designed accuracy is 20 m horizontally and 16 m vertically. To minimize the influence of vegetation, attempts were made to obtain the final products from data for leaf-off periods. The SRTM elevations used in this study are the second version of the SRTM data, at a resolution of 3 arc second (often quoted as 90 m resolution), which exhibit well-defined water bodies and coastlines, and the absence of spikes and wells (Farr et al. 2007).

Niche algorithm

Maximum Entropy (MaxEnt) was selected as the niche algorithm for this study, which is one of the most popular ENM algorithms currently available (Elith et al. 2006). The environmental variables used in this study are precipitation, maximum temperature, minimum temperature, and mean temperature. To capture the seasonality across our study area and limit the number of climate surfaces, we used four out of the twelve months (January, April, June, and October) (McPherson and Weltzin 2000). For topography, we use the commonly used elevation layers (DEM), obtained from United States Geological Survey (USGS) (Farr et al. 2007), having a spatial resolution of 90 meters, and since all of the climate surfaces are at 1 km, the DEM was upscaled to 1 km using mean.

Creating uncertainty layers/points

To evaluate the influence of the uncertainty from different inputs on the niche modeling result, eight different types of models were included in the analysis (Table 2):

- (M0) No uncertainty (base model)
- (M1) Uncertainty on climate variables
- (M2) Uncertainty on climate variables and DEM
- (M3) Uncertainty on climate variables and point locality
- (M4) Uncertainty on all (climate variables, DEM, and point locality)
- (M5) Uncertainty on DEM
- (M6) Uncertainty on DEM and point locality
- (M7) Uncertainty on point locality

Table 2: Description of each model uncertainty parameters.

Model Number	Climate	DEM	Point
M0	No	No	No
M1	Yes	No	No
M2	Yes	Yes	No
M3	Yes	No	Yes
M4	Yes	Yes	Yes
M5	No	Yes	No
M6	No	Yes	Yes
M7	No	No	Yes

For each model, we hypothesized that the uncertainty brought by the point locality, climate surfaces and topographic data are independent and random from each other (Heuvelink et al. 1999), and a numeric error propagation model was therefore implemented,

$$\varepsilon_{ENM} = \sqrt{a\varepsilon_{locality}^2 + b\varepsilon_{ClimSurf}^2 + c\varepsilon_{topography}^2}\ , \tag{8}$$

where ε_{ENM} is the uncertainty of niche modeling result, $\varepsilon_{locality}$, $\varepsilon_{ClimSurf}$, and $\varepsilon_{topography}$ are uncertainties brought by point locality, climate surfaces and topography inputs, respectively, and a, b, and c represent the weight of the corresponding input (if the variable was included in the model, it was set as 1, otherwise 0).

The point locality uncertainty was introduced by using an uncertainty field model (Guo et al., 2008). To generate uncertainty in the point localities for each species, we created a buffer around each locality; the distance was obtained from the uncertainty given by GBIF. Then, randomly, 100 different sets of points were generated for each species. This gave us 100 unique localities for each of 43 species (total of 4,300 set of points).

To generate both the climate and topography layers that contain uncertainty, we ran a Monte Carlo simulation of 100 per surface. Utilizing the standard deviation and mean of the error (Table 2), we were able to run an inverse cumulative density function and obtain a random value every time each was added to the original (9),

$$L_1 = L_0 + \in, \tag{9}$$

where L_1 is the new layer containing uncertainty, L_0 is the original layer without uncertainty and $\in$ is the random number obtained from the inverse cumulative density function. This was done for all of the four different climate variables (total precipitation, maximum, minimum, and mean temperature) for all four months (January, April, July, and October) and for DEM (Table 2).

Table 3: Standard deviation and mean error for each variable and month.

	Precipitation		Temp Mean		Temp Min		Temp Max		DEM	
	STDV	MEAN	STDV	MEAN	STDV	MEAN	STDV	MEAN	STDV	MEAN
Jan	52.5	-0.045	4.1	0.010	4.8	0.020	3.4	0.006	NA	NA
Apr	36.2	-0.016	4.1	-0.003	3.3	0.004	3.3	-0.002	NA	NA
Jul	54.7	0.045	4.6	0.005	4.9	0.006	4.7	-0.002	NA	NA
Oct	41.2	-0.002	5.4	-0.015	5.5	0.006	5.6	-0.013	NA	NA
Fix	NA	NA	NA	NA	NA	NA	NA	NA	4.0	0.100

The point localities were divided into three random sets: 60% for training the model, 20% for validation, and 20% for testing. Each set of models will be run 100 times for each species. For each of the model output, a conversion will be made from continuous to binary utilizing a 5% omission rate as the threshold (Pearson et al. 2004, Li and Guo 2010). This is necessary to determine how much each model has changed. The validation points are used to extract the model values, which are then sorted from lowest to highest and the 5% lowest value will be accepted as the threshold.

Uncertainty analysis

To test how different input uncertainty influence the niche modeling result, the McNemar's test, a widely used statistical test for paired nominal data, is selected to determine the change compared to our base model (M0)(McNemar 1947). By applying a 2×2 contingency table with a dichotomous trait, the existence of marginal homogeneity (i.e., the row and column marginal frequencies are equal) is tested under chi-square distribution. In this study, the null hypothesis is that the different uncertainty models cannot bring significant change in ENM modeling results. The test statistic with Yates's correction for continuity is constructed as follows (Yates 1934),

$$\chi^2 = \frac{(|P_b - P_c| - 0.5)^2}{P_b + P_c}, \tag{10}$$

where P_b is the number of presences predicted from M0, but showing as absence in uncertainty model; P_c is the number of absences predicted form both M0 but showing as presence in uncertainty model. The χ^2 should conform to a chi-squared distribution when P_b and P_c are sufficiently large. If the possibility of the calculated $\chi^2$is smaller than the pre-

defined significance level (α=0.01), the null hypothesis will be rejected; otherwise, it will be accepted.

Moreover, the Kappa coefficient (k), which can reflect the inter-rater agreement of categorical items, was selected to represent the similarity of niche modeling result considering uncertainties to the base model (M0). The Kappa coefficient was calculated from the following equation,

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}, \tag{11}$$

where Pr(a) is the relative observed agreement, and Pr(e) is the hypothetical probability of chance of random agreement.

Besides, we also calculated the proxy of F-measure based on positive-background data (Fpb) which is defined as follows:

$$F_{pb} = \frac{2*TP}{TP + FN + FP}, \tag{12}$$

where TP is true-positive, FN is false-negative, and FP is false-positive

In this study, Kappa was calculated in two different ways. One is where the entire model study area (North America) was used, meaning that all the cells were involved in the model. The second was based on the test set of points saved in the splitting of the data; those points, plus an additional 500 random points per species, were used to calculate Kappa and Fpb.

RESULTS

Comparison by all species

The pattern of the two different ways in which Kappa was calculated did not change the outcome of the analysis. Both ways (Figure 4, A and B) suggest M5 (uncertainty on DEM)

was closest to the base model (M0). The results suggest based on this study, DEM has the least uncertainty effect on ENM distribution. The most different models are the ones that consider climate uncertainty; from M1 to M4 across both Kappa's they are the lowest. The lowest is M4 (uncertainty on all), however, the lowest mean is M3 (uncertainty on climate and points). The comparisons from Kappa and Fpb (Figure 4 B and C), are almost identical, but Fpb's scale is almost twice as big as Kappa's. In general, all three sets of comparisons are constantly displaying the same general trend. The models are different from the base and it is statistically significant different (Table 4). This is across all models, where the P value for all models is < .05. After a closer look at Appendix B, Figures 1-8, it becomes more evident that the models are different compared to the base (Appendix B Figure 1). A model with less uncertainty (like M5) is more close to the base model, and has higher predictive accuracy. Accordingly, the base model would have the highest accuracy, as it does not consider any uncertainty. From this perspective, a model that considers uncertainty will decrease the predictive accuracy, as it does not consider any uncertainty.

Table 4: McNemar for each species, results are based on test set point plus 500 random points.

Species	Group	P (M1,M2,M3, M4,M5,M6,M7)	Species	Group	P (M1,M2,M3, M4,M5,M6,M7)	Species	Group	P (M1,M2,M3, M4,M5,M6,M7)
2436042	Broad	* , , , , , ,	2438609	Broad	* , , , , , ,	6163387	Broad	* , , , , , ,
2436078	Broad	* , , , , , ,	2438621	Broad	* , , , , , ,	6163394	Broad	* , , , , , ,
2437427	Narrow	* , , , , , ,	2439141	Moderate	* , , , , , ,	6163501	Broad	* , , , , , ,
2437867	Narrow	* , , , , , ,	2439473	Broad	* , , , , , ,	6164308	Broad	* , , , , , ,
2437874	Broad	* , , , , , ,	4263215	Narrow	* , , , , , ,	6164672	Broad	* , , , , , ,
2437967	Broad	* , , , , , ,	4263341	Broad	* , , , , , ,	7059247	Moderate	* , , , , , ,
2437981	Broad	* , , , , , ,	4263615	Moderate	* , , , , , ,	7194076	Broad	* , , , , , ,
2437985	Broad	* , , , , , ,	4263628	Broad	* , , , , , ,	7194090	Moderate	* , , , , , ,
2438021	Broad	* , , , , , ,	4263661	Broad	* , , , , , ,	7194092	Moderate	* , , , , , ,
2438028	Moderate	* , , , , , ,	4265036	Broad	* , , , , , ,	7261500	Broad	* , , , , , ,
2438038	Broad	* , , , , , ,	4827935	Broad	* , , , , , ,	7261506	Narrow	* , , , , , ,
2438438	Broad	* , , , , , ,	5219681	Broad	* , , , , , ,	7261509	Broad	* , , , , , ,
2438451	Narrow	* , , , , , ,	5219683	Broad	* , , , , , ,	7261539	Broad	* , , , , , ,
2438516	Broad	* , , , , , ,	6163341	Broad	* , , , , , ,			
2438603	Moderate	* , , , , , ,	6163386	Broad	* , , , , , ,			(* < .05)

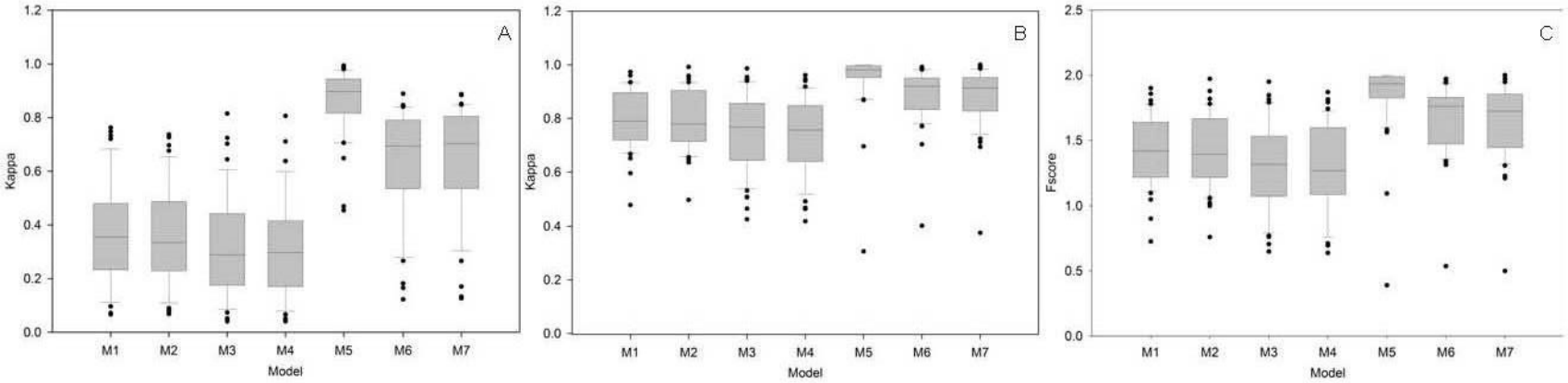


Figure 4: All three graphs (A/B/C) display the entire species list, for (A/B) it displays the Kappa range as for (C) it displays the Fpb range (y- axis). (A) Uses the entire model area, meaning all the cells in the image/model, (B/C) uses the points save in the beginning (testing points) plus an additional 500 random points. For the x-axis, please refer to Table 3

Comparison by species group

There are very little differences from separating the analysis into groups (Figure 5). Across all three groups (narrow, moderate, and broad), each group and type of test consistently show M5 being the most similar to the base model and M3/M4 as the least similar. The narrow group (Figure 5, row 1) shows that the Kappa/Fpb values have a bigger range than the other groups; for example, M6 for Figure 5 A1 ranges from .4 to .8 while the other maximum range is around .2. For the moderate group (Figure 5, row 2), notice that Kappa (Figure 5, B2) are all around .7 to 1 and this is the only set that displays a cluster of all the models. However, even with that clustering, again, M5 is the closest to the base and M3/M4 are the most different from the base. For the broad group (Figure 5, row 3), the results are identical, since this is also due to the fact that most of the species are classified as broad. The variation, however, shows that M5 species are all clustered, while the other models have a bigger variation.

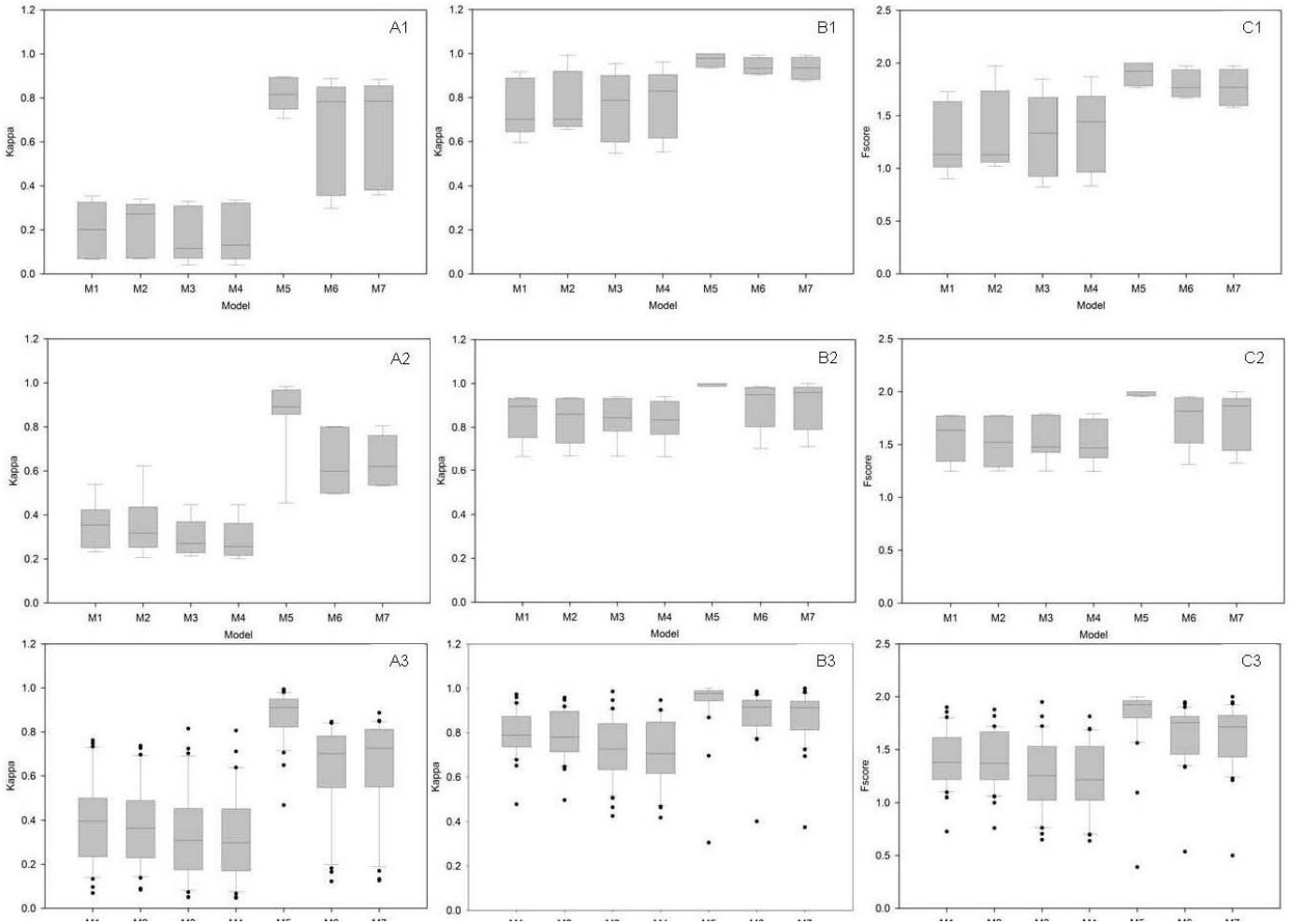


Figure 5: All (1)s are the entire range for narrow classified species, (2)s are moderate classified species and (3)s are for broad classified species. For (A/B) it displays the Kappa range as for (C) it displays the Fpb range (y- axis). (A) Uses the entire model area, meaning all the cells in the image/model, (B/C) uses the points save in the beginning (testing points) plus an additional 500 random points. For the x-axis, please refer to Table 3

Comparison by individual species

Almost every species shows the same pattern as previously seen (Figure 6). Looking at the comparison of Kappa/Fpb by species and group we do, however, see that some species are somewhat different than the trend. For the narrow species (Figure 6, row 1), M5 to M7 are all the highest for each test (Kappa/Fpb), where those models are the ones without climate uncertainty. This pattern is clearly seen in panel A1 (Figure 4) where M1 to M4 are all between 0 to .4 while the other models, M5 to M7, are all between .4 to 1. For the moderate species (Figure 6, row 2), the trend is the same as mentioned above. The interesting model is M5 for panel B2 and C2, as it seems constant across all species and has the highest Kappa/Fpb from the rest. For broad species (Figure 6, row 3), the pattern is still consistent but with variation; since it is the group with the most species, it is somewhat harder to see. Models M5 to M7 still tend to be in the upper range while M1 to M4 are towards the bottom. However, for at least one species, 5219681 (Sciurus Carolinensis Gmelin, Squireel), that observation is reversed. M5 to M7 are in the bottom range while M1 to M4 are on the top range (Figure 6, B3 and C3).

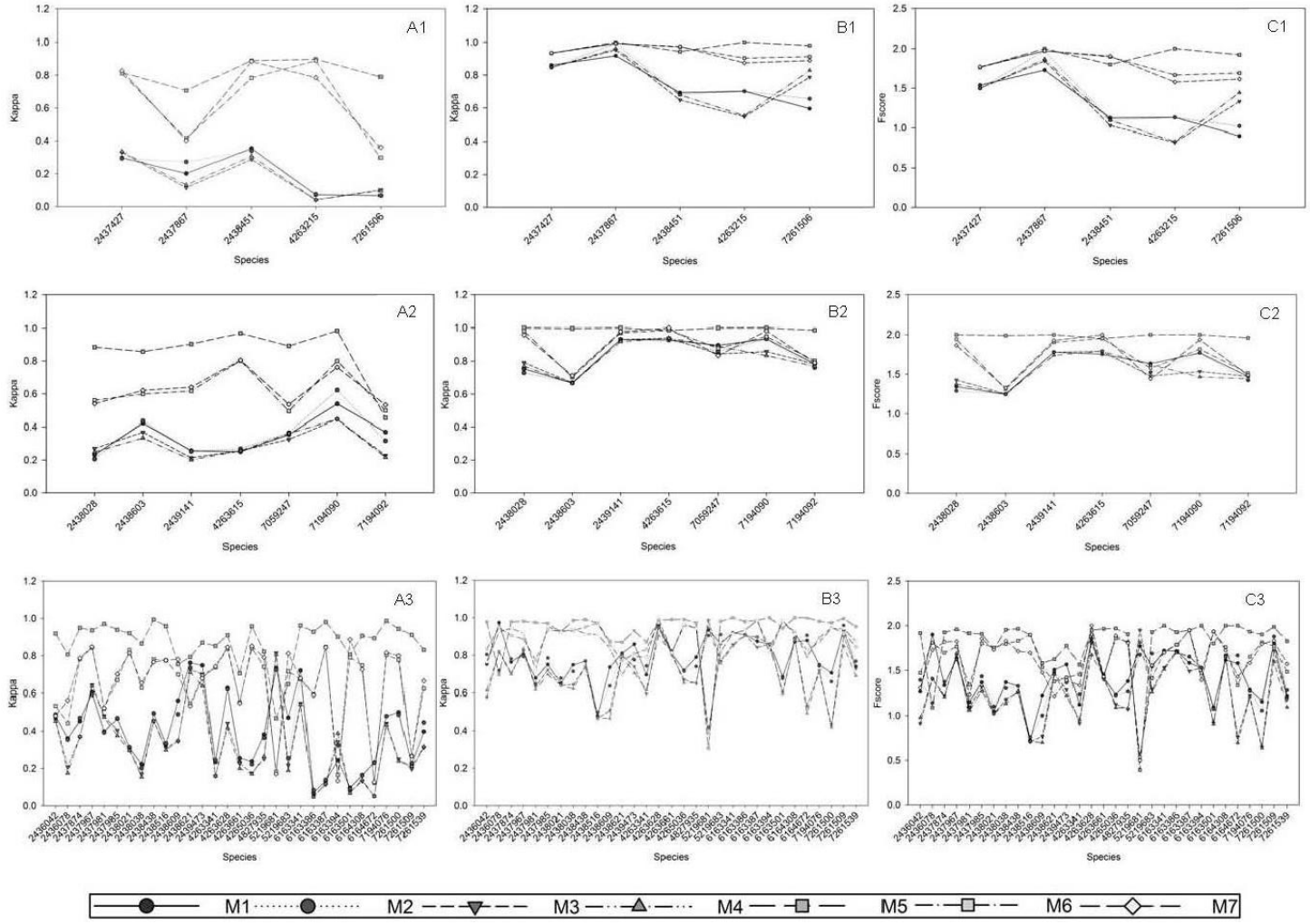


Figure 6: All (1)s are for narrow classified species, (2)s are moderate classified species and (3)s are for broad classified species. For (A/B) it displays the Kappa range as for (C) it displays the Fpb range (y- axis). (A) Uses the entire model area, meaning all the cells in the image/model, (B/C) uses the points save in the beginning (testing points) plus an additional 500 random points. For the x-axis, please refer to Table 1, Appendix B.

DISCUSSION

Impact of uncertainty

The impact of uncertainty in all three sources does have an effect on the outcome of the predictions and it is statistically significant (Table 4). Based on the statistical analysis done on each species/group, results suggest (Figures 4 – 6) that each model that includes any type of climate uncertainty is more different from the ones that do not. It also implies that DEM and point uncertainty are closer to each other while being close to the base model. Based on these results emphasis should be given in trying to understand how the climate surfaces are generated, since it will affect the model output. This knowledge is critical for the understanding of the outputs of the models and the error associated with each location. It can also state that for this study, the error propagation for all three sources should be considered in any type of niche models since each model had a significant difference from the base model, as seen in somewhat similar studies (Kriticos and Leriche 2010, Beale and Lennon 2012, Fernández et al. 2013). Since the most common approach in niche modeling is to only use variables that are related to climate and limited of topography. One thing to consider is that variation for model M5 is the least from the rest of the models, as mentioned in the results section. This could be an indicator that DEM uncertainty alone is very similar no matter the species/group and closest to M0. While one might make the assumption that since the model only consists of one type of uncertainty, that might be the reason, however, models M1 and M7 all have one type of uncertainty (climate or point) and those show that the variation is quite large.

Classification comparison

The analysis was conducted in multiple ways to try to determine if there were any cases in which some species might react differently than another. As shown in Figures 4-6, there were not many overall differences. Mainly, the differences where either M3 or M4 being the most significant compared to the base and the range values of the statistical test. However, the range of values would have something to do with the amount of species in each group. In addition, one key observation is that M5, across all types of analysis, has proven to be the most alike to the base; M5 considers only uncertainty in DEM. These findings can suggest that for at least the 43 mammalian slope is an important factor, since when running the Monte Carlo simulation the elevation was randomly changed, but the slope stayed the same. As it has been widely studied, elevation gradient are well known to have restriction on the species potential habitat (Austin 2007, Sexton et al. 2009). Both Kappa and Fpb patterns are very similar, with the only differences being the y axis. One possible reason is that, the background dataset only contains a small proportion of presence

data, and their influence is small. So the "pseudo" absence (background) data will act like real absence data. In this way, Fpb and kappa will perform similarly. Also as mentioned above, models M5 constantly show the least variation from the rest; while this pattern generally stands, there are cases where it does not (Figure 5, A2). Models M6 and M7 variation is also clustered together compared to the models that do have climate uncertainty.

Improvements and conclusion

There are several questions and additional studies that would improve our understating of how uncertainty affects ENM. One aspect which the study did not take into consideration was running multiple niche algorithms. As mentioned, there are many well establish algorithms in our community. This is another dimension of how a species distribution might change. The study was also limited to Mammalians, therefore, additional research should be done on other types of species. For example, in Oak Trees and Redwoods, would these species model distribution change based on uncertainty, or since these species take longer to “move” would it not matter? One other factor which more research should be done on, is in the fact that only climate and DEM were used in this study; the use of more appropriate variables could affect the outcome. As mention, DEM uncertainty was the closest to the base model and this can be attributed on the restriction of the species, therefore, other species that do not have these restrictions should be analysis.

In conclusion, we found that generating the models with different types of uncertainty does have a very significant effect on the outcome of each model. Climate uncertainty should be considered when modeling a species, since based on the climate product used the results might be significantly different, which will lead to false judgment. There is still much research to be conducted to fully understand the impact of error in each parameter and the effect it has on the outcome.

CHAPETER 5: THE IMPACT OF UTILIZING BIOPHYSICAL VARIABLES DERIVED FROM LIDAR TO MODEL THE PACIFIC FISHER HABITAT IN THE SIERRA NATIONAL FOREST

INTRODUCTION

Ecological niche modeling (ENM) is a tool which provides a probability of distribution for a species based on the parameters that the model is being fed (Feria and Peterson 2002, Graham et al. 2004, Guo et al. 2005, Rodriguez et al. 2007, Peterson and Nakazawa 2008, Lenoir et al. 2010, Fernández et al. 2012). ENM have also been used to predict other scenarios, such as classifying land cover to remote sensing images and determining suitable helicopter landing areas (Li and Guo 2010, Li et al. 2011, Doherty et al. 2012). Basically, ENM builds a statistical relationship between the occurrences of the species to environmental predictors, allowing it to predict future/current potential suitable habitat under any type of environment, or such layers. Unfortunately, most models are only given climatic variables and some biophysical variables (e.g. elevation, slope, aspect). This is a good start; however, some species depend on other biophysical variables because of their habitat constraints.

The Pacific Fisher (*MartesPennanti*), a medium-size mammal historically distributed from the Boreal forests of Canada through the northern United States and into the Sierra Nevada Mountains, has recently declined dramatically in range and abundance. Commercial over-trapping, change in forest structure associated with logging and altered fire regimes, increased human access, and habitat loss to urban and recreational development are the primary reasons for the decrease in range and abundance of the Fisher (Ruggiero et al. 1994, Zielinski et al. 2005, Purcell et al. 2009, Zielinski et al. 2010). The West Coast Fisher populations have been petitioned for listing under the Federal Endangered Species Act on three occasions, but these petitions were denied, partly because of lack of empirical information (Purcell et al. 2009). As a result, larger efforts to monitor and assess its habitat associations are urgently needed to identify required habitat conditions and to evaluate the potential impacts of habitat change on Fisher populations (Zielinski et al. 2010).

Lidar recently has emerged as an important optical remote sensing technology in the past decade. It has become popular due to its ability to generate 3D data with high spatial resolution and accuracy. Lidar works by measuring the properties of scattered light to find the range, or other information, of a distant object (ESRI 2009). Basically, it finds the range of the object by measuring the time delay between the transmission's pulse (signal sent by the laser) and the time the signal is received. Lidar has been used extensively in creating 3D urban models which has been a great tool in capturing many critical features that normally are not caught by traditional remote sensing products (Robine et al. 2008, Alvarez et al. 2013). Currently, Lidar technology is being used in many environmental fields such as: geology, ecology, climate change, geography, geomorphology, and seismology (Brooks and Doswell 1993, Wieczorek et al. 2004). However, there is a huge difference between urban and forestry structures. Urban areas are usually structured in such a way that not much information is needed (point density) to be able to capture objects, whereas

the environment (forestry, agriculture, etc.) is much more complicated since nothing is structured and thus, requires more information to make a better model. This leads to 3D modeling of the environment being much more costly and computationally expensive. Lidar has shown promise to derive a variety of forest biophysical parameters, including Digital Surface Models (DSM), tree height, diameter at breast height (DBH), basal area, canopy density, wood volume, biomass, leaf area index (LAI), fraction of photosynthetically active radiation absorbed by the canopy (fPAR), crown diameter, stand density, carbon stocks, and etc. (Wieczorek et al. 2004, Wang et al. 2011)

In general, there is a lack of studies trying to determine the influence of inclusion or exclusion of biophysical variables in ENM. In terms of the Pacific Fisher, a species which depends mostly on forest structure, it would seem that biophysical variables should have higher importance than the normal variables included in ENM. We propose the question: would adding or replacing biophysical variables outperform the generic/standard approach? The answer to this question is essential to better understand and predict the potential suitable habitat of an endangered species like the Pacific Fisher. Multiple statistical tests will be run to determine the difference and the significance on comparing biophysical, physical, and the combination of both.

METHOD

The study areas are located in the northeast region of Oakhurst, California, named Sugar Pine. This area is part of SNAMP project which covers about 36 km^2 (Figure 7). This area is topographically complex with elevations ranging from 758 m to 2,652 m. Primary tree species in the order of abundance include *Calocedrus decurrens* (Incense Cedar), *Abies concolor* (White Fir), *Pinus ponderosa* (Ponderosa Pine), *Pinus lambertiana* (Sugar Pine), *Sequoiadendro giganteum* (Giant sequoia), *Quercus kelloggii* (Black Oak), *Quercus spp.* (Live oak), *Cornus nuttallii* (Mountain Dogwood), and *Ainus rhombifolia* (White Alder). The composition of these primary species shows a general pattern for all plots as a whole, and within each single plot, species composition among plots may vary significantly.

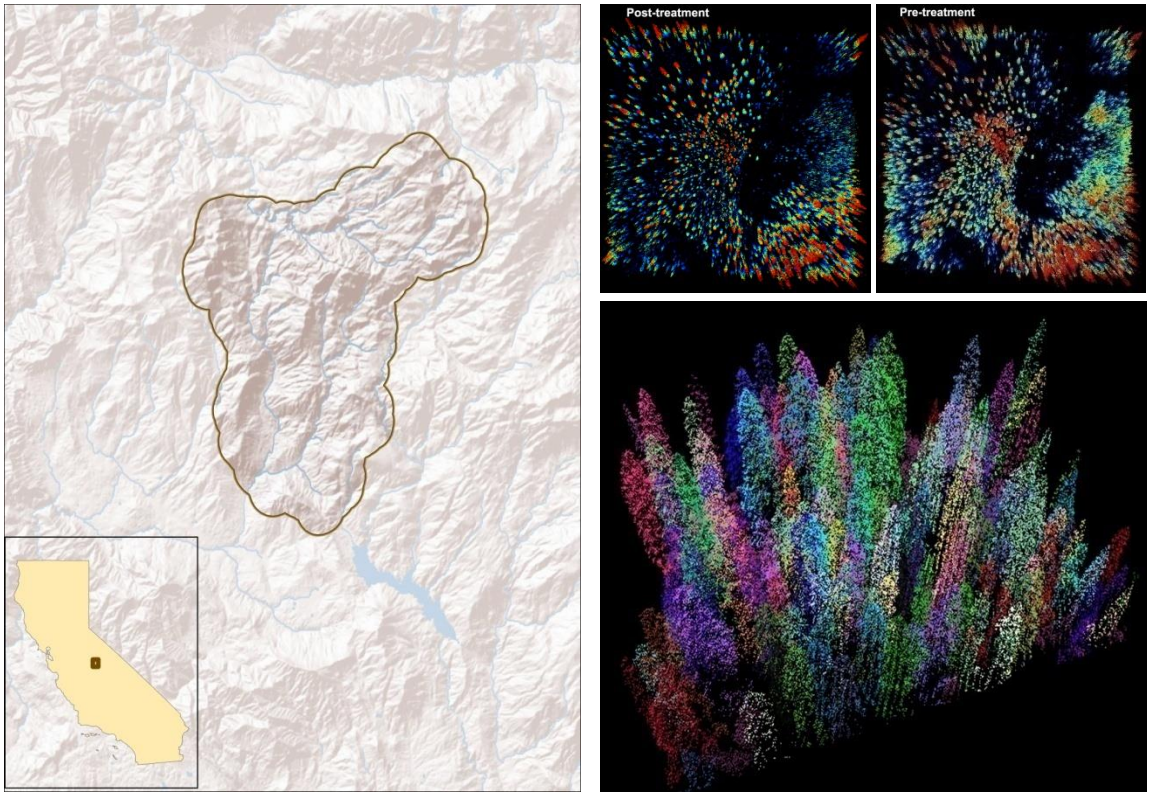


Figure 7: (Left) Sugar Pine, SNAMP study area, (Top right) Lidar displaying pre and post thinning treatment, (Bottom right) Lidar detection of tree segmentation.

Since Lidar is in meter resolution, a new subset of climate data were generated using the methods of ClimSurf (Alvarez et al. 2013). ClimSurf was created using observational weather stations from Food and Agriculture Organization (FAO) and the Global Historical Climate Network (GHCN) version 2 (Peterson and Vose 1997, FAO 2001) using, Thin Plate Spline (TPS) which is highly used in creating climate surfaces (Hutchinson and Gessler 1994, Hutchinson 1995, Price et al. 2000, Jeffrey et al. 2001, Hijmans et al. 2005, Hong et al. 2005, Hancock and Hutchinson 2006, Hancock and Hutchinson 2006, Tait et al. 2006). Since ClimSurf was generated at 1 km, the clean weather station was obtained to generate a new product at 20 meters. The weather station included 552 stations for precipitation, 357 stations for maximum temperature, 381 stations for minimum temperature, and 415 stations for mean temperature. This data were cleaned manually by graphing each station for yearly/monthly and determining if any outliers have been detected. Since Precipitation does not follow any pattern, a different procedure was applied: the outliers were detected through the use of a spatial process of including all the information surrounding the target station. The weather station average from 1950 to 1999 for all four variables, and a DEM at 20 meter resolution was used as covariate for temperature and a combination of Radar was used for precipitation since in Alvarez et al. (2013) it was found that it acts as the best covariate.

Lidar data are acquired by the National Center of Airborne Laser Mapping (NCALM) at the University of Florida. The surveys use an Optech GEMINI Airborne Laser Terrain Mapper (ALTM) mounted in a twin-engine Cessna Skymaster (Tail Number N337P). The study area was covered in five survey flights: two each on September 13 and 14, and one final flight on September 15, 2007. To be able to validate the data, the site was visited multiple times by members of the GIS lab at UC Merced and Kelly's lab at UC Berkeley. The ground-truth data were collected to systematically calibrate and validate the Lidar derived products, and be correlated with Lidar data to upscale site measurements to the landscape level. The sampling design is based on Jensen et al., 2008 and Paletto et al., 2009. Around 30-40 sample plots are distributed throughout the entire study area with a radius of 15 m. In this study, the biophysical variables are: Mean Height, Max Height, Height to Live Canopy Base and Diameter at Breast Height, Crown Radius, Canopy Cover, and Leaf Area index. They are generated based on the following:

All vegetation variables are generated by using a regression based approach, by first extracting a subset of the raw Lidar data in each plot location. Then the points are normalized by subtracting the ground points (DEM), while a height profile is created on the normalized points using the following groups: z values for minimum, percentiles (1^{st}, 5^{th}, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th}, 95^{th}, 99^{th}), maximum, mean, standard deviations and the coefficient of variation. The last step is to apply the best model base on the stepwise regression model in which the Lidar-base predictors are fitted against the field measurements.

Canopy Cover (CC) is generated by first analyzing the canopy height model (CHM) which is at 1 m resolution. The value of the canopy cover pixel is calculated as the ratio of CHM pixels that have a value above a threshold to the total number of extracted pixels from the (Lucas et al. 2006).

The leaf area index (LAI) variable is created using the Lidar vegetation points, normalized by the DEM. An average scan angle is calculated using the extracted Lidar points and the following equation:

$$a = \frac{\sum_{i=1}^{n} angle_i}{n} \tag{13}$$

where a is the average scan angle, n is the number of extracted points and $pt_n scan\ angle$ is the scan angle for a single extracted point i. Next the gap fraction (GF) is calculated using the following equation:

$$GF = \frac{n_{ground}}{n} \tag{14}$$

where n_{ground} is the number of extracted points that have a z value smaller than 1.5 m (equivalent to the height of a hemispherical camera) and n is the total number of extracted points. Finally, the LAI value is calculated using the following equation:

$$LAI = -\frac{\cos(ang) \times \ln(GF)}{k} \tag{15}$$

where LAI is the extinction coefficient and ln is the natural logarithm (Richardson et al. 2009). The value 0.5 is used for the extinction coefficient k, as suggested in the literature (Richardson et al. 2009). The tree segmentation product was thinned by randomly removing points at 5 different levels: 10%, 20%, 30%, 40% and 50%. At each level, the segmentation product was reduced by the corresponding percentage (i.e. 10% means 10 percent of the points were removed from the segmentation product).

Maximum Entropy is one of the most popular niche algorithms currently available (Elith et al. 2006) and is the one being used in this study. To capture the seasonality across our study area and limit the number of variables, only four months have been selected (January, April, June, and October) (McPherson and Weltzin 2000). The Fisher data consist of 6 857 unique localities, meaning that only one location per 20 meters was kept. The data were than divided into three random sets: 60% for training the model, 20% for validation, and 20% for testing. Since the statistical test being conducted for this study only can handle binary outputs, the 5% omission rate as the threshold (Pearson et al. 2004, Li and Guo 2010) will be applied to each model output. The validation points are used to extract the original continuous model values, those values are then sorted from lowest to highest and the 5% lowest value will be accepted as the threshold.

To determine how different each model is from another, two different measurements will be applied, where one is the very well-known Kappa (k). Kappa coefficient was calculated from the following equation,

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(\mathrm{e})} \tag{16}$$

where Pr(a) is the relative observed agreement, and Pr(e) is the hypothetical probability of chance of random agreement. The other measurement also calculates the proxy of F-measure based on positive-background data (Fpb) which is defined as follows:

$$F_{pb} = \frac{2*TP}{TP + FN + FP} \tag{17}$$

where TP is true-positive, FN is false-negative, and FP is false-positive. To determine the significance level, McNemar's test was applied. McNemar is a widely used statistical test for paired nominal data, selected to determine the change compared to each set of models (McNemar 1947). By applying a 2×2 contingency table with a dichotomous trait, the existence of marginal homogeneity is tested under chi-square distribution.

$$\chi^2 = \frac{(|P_b - P_c| - 0.5)^2}{P_b + P_c} \tag{18}$$

where P_b is the number of presences predicted from M_1 and absence in M_2; P_c is the number of absences predicted in M_1 but showing as presence inM_2, where M is any model output. The χ^2 should conform to a chi-squared distribution when P_b and P_c are sufficiently large. If the possibility of the calculated $\chi^2$is smaller than the pre-defined significance level (α=0.01), the null hypothesis will be rejected; otherwise, it will be accepted.

RESULTS

Influence of biophysical variables

The similarities between using all the variables and using either subset are statistically significant (Table 5 and Table 6) no matter how McNemar is analyzed. Comparing all the different types of models to each are significantly based on our results of Kappa and McNemar (Table 5), in addition, results suggest that the comparison between the physical model and the all-model Kappa is high (~0.9), meaning that they are very similar. In comparison with using all the variables versus biophysical model results suggest that they are not similar having a Kappa value being very low (~0.36) with a p-value of < .05; these patterns also are similar to the comparison of physical to the biophysical model. Respectively having a Kappa value of (~0.3) and a p-value of < 0.05.

Table 5: Base model results utilizing the full model output.

	Physical		Biophysical	
	Kappa	P	Kappa	P
All	0.9261	< .05	0.3684	< .05
Physical	NA	NA	0.3189	< .05

Table 6: Base model results utilizing the test and background points.

	Physical			Biophysical		
	Kappa	P	Fpb	Kappa	P	Fpb
All	0.8729	< .05	1.9739	0.2949	< .05	1.8509
Physical	NA	NA	NA	0.2317	< .05	1.8351

Comparing the model output visually, the difference suggest not to be noticeable versus comparing at the all and physical model (Figure 8). The main difference suggest that the physical model has a smooth affect while the all-model has a "patchy". In comparing the two models that contain physical variables to biophysical model, the difference are visually noticeable. The biophysical model covers the whole region, but having the "patchy" effect.

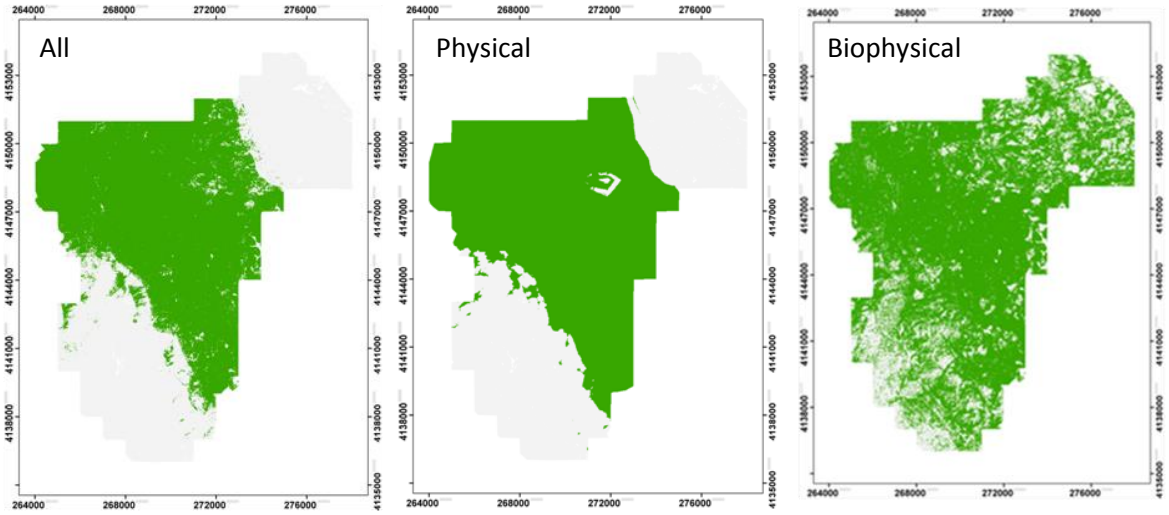


Figure 8: Base model comparison

Projecting based on forest thinning

The projection of the Fisher on forest thinning is very interesting; notice that Kappa does decrease based on the percentage of thinning (Table 7). At 10% thinning projection, utilizing all of the variables, Kappa is ~0.99, while by increasing the thinning projection Kappa keeps getting smaller, down to ~0.93 (30% thinning). While the magnitude on the all-variable model decreases, it does so at a slow rate. This is different from the biophysical model. Here, the projections of 10% Kappa are ~0.93, but differ from the all-model at 20% thinning where the biophysical model is ~0.86 and at 30% it is ~0.77. This reduction of Kappa is ~25% less than the base 10% projection, meaning that the forest thinning is having an effect on the habitat potential niche. Visually, the differences between the all and biophysical models for each thinning percentage are well noticeable (Appendix C Figure

1); however, the differences for the biophysical or all-models based on all of the percentages are not different. There are minor "patchy" differences, but they are not visually different.

Table 7: comparing to multiple forest thinning scenarios.

	All			Biophysical		
	Kappa	P	Fpb	Kappa	P	Fpb
All	0.9923	< .05	1.9985	NA	NA	NA
Biophysical	NA	NA	NA	0.9320	< .05	1.9845

Base to 10 % comparing

	All			Biophysical		
	Kappa	P	Fpb	Kappa	P	Fpb
All	0.9695	< .05	1.9938	NA	NA	NA
Biophysical	NA	NA	NA	0.8656	< .05	1.9691

Base to 20 % comparing

	All			Biophysical		
	Kappa	P	Fpb	Kappa	P	Fpb
All	0.9398	< .05	1.9877	NA	NA	NA
Biophysical	NA	NA	NA	0.7715	< .05	1.9493

Base to 30 % comparing

DISCUSSION

Importance of biophysical and physical parameters

As shown in Figure 8, the differences between using physical and biophysical variables are not only visually different, but also statistically different (Tables 5-6). Utilizing biophysical variables only, suggest to have more of a detailed and broad effect on the model, in comparison to utilizing physical variables. When combining all of the variables, our results suggest that the physical variables have a greater influence on the range of potential habitat of the species. This observation also holds true when comparing the Kappa values (Table 6), as previously mentioned, in that the comparison of the all variables model to physical variables is high compared to the value of the biophysical model, which is very low, a difference of .7. With this, one can state that physical and the all-model are very similar to each other, while there is a significant difference with the biophysical model. Once again,

the Fpb value follows the same pattern as Kappa, having the comparison of physical to the all-model reveal that the only difference is that the magnitude or percentage of difference on the value is not big.

Forest thinning projection

As mentioned in the results, the comparison between using all-variables and just the biophysical variables is, indeed, visually and statistically different. The biophysical model covers the whole study area, but displays somewhat more detailed and the potential suitable habitat for the Fisher. While using all of the variables, the model suggest to be constrained by one of the variables, this variable being April's precipitation, which has about 30% influence on the model. More interestingly, the top five influential variables are all precipitation variables: April, July, January, and October (in order of importance). The second and only different influential variable is maximum temperature, at 20%. No matter the percentage of thinning projection, this trend remains intact with some of the bottom top five variables swapping in influence on the model. However, while they might be visually equal compared to the current model, Kappa seems to decrease with comparison to the base.

Future considerations and improvements

Most ENM only have available a few point locations and the main problem is in trying to determine the niches for species without a reasonable sample size (Pearson et al. 2007). However, for our study, we found that the model was fed too many points. The total number of unique points are ~6 900 and 4 114, wherein use to train the model there may be a problem in which the model was given too many points for that small region. As seen in Figures 1-2, visually all of the areas that are predicted as suitable for the Fisher appear to follow the pattern of the location of the species. One way to determine if the model has been over-fitted is to run the model with different amounts of point density. This way it can be proven if, or not, the model is being over-fitted with point localities.

The model took input for the location of the Fisher, where one might make the assumption that, for at least the Fisher, a more realistic approach on determining the best suitable habitat is to use the location of their nest. The denning structures are considered to be the most important habitat components for fishers; they exhibit the greatest selection for natal dens and resting sites (Gruber and Levizzani 2008). Other studies have shown the capability in determining resting sites, suggest that these are structures used as protection from predators and inclement weather and as a way to regulate body temperature (Ruggiero et al. 1994). In addition, studies have found that the areas in which these Fisher resting sites were located had higher canopy cover, large trees and snags, and structural complexity. Abiotic features such as distance to water, slope position and steepness also have shown to be indicative of areas supporting fisher resting structures (Zielinski et al. 2005). Fishers use natal dens and maternal dens to reproduce and raise their kits, and suitable dens are critical to the Fisher population development. Modeling the Fisher with the denning site

might be a better representation of the potential habitat of this very critical endangered species.

Another key factor is to include other such parameters that would help improve the model, such as hazard areas, obstacle/road block, and even simple data such as closest water supplies. These will potentially have somewhat of an influence on the model (Zhao et al. 2012). Therefore, more research needs to be conducted not only on the Fisher, but in general to determine which parameters have the most influence and impact on ENM and on which group of species, if any. Also, the forest thinning method that was applied on the Lidar data can be improved and there should be more research dedicated to finding out which method realistically can be applied.

In conclusion, the study shows that utilizing biophysical or physical variables does affect the output of the ENM. The differences for all three model types are significant different and as well as visually. Our analysis based on projections of forest thinning shows that the fisher habitat will change and the change is significant. Our study demonstrate the influence of which variables are included in model can and will affect the either model, therefore, one must be selective and analyze which set of variables are included.

CHAPTER 6: CONCLUSION & FUTURE WORK

The traditional method in which DEM is utilized as a covariate for all the variables has been proven to be insufficient, at least for precipitation. Significant work has been done to determine what covariates are best suited for each climate variable. I have shown that there are better products which will help reduce the uncertainty, however, there might be other products that are even better. Radar has been proven to be the best covariate for precipitation since precipitation is one of the variables that tends to have the most uncertainty. Even with the improvements there are still limitations in utilizing these climate surfaces. Generating climate surfaces of a finer spatial resolution is (1) time consuming and (2) might produce unusable data. Results suggest that the source dataset utilized in creating these climate surfaces have a problem with the number of decimal places it contains in its longitude and latitude fields. This is an example of the limitations that need to be considered when creating climate surfaces.

Uncertainty in ecological niche modelling has been a hot topic in the environmental community, and we have made significant work in determining which source of uncertainty in the input data has the most influence when modeling a species. Results suggest that any type of uncertainty in modelling a species is significant and climate uncertainty has the most influence compared to the base model. Results suggest that adding more meaningful variables does have an impact on the size of the modeled potential habitat. Biophysical and physical models are significantly different; therefore, one needs to select variables to be included into the model carefully to produce more realistic results.

Overall, the study was successful in incorporating remote sensing data to reduce uncertainty when creating climate surfaces. Even though there are still limitations when creating very fine spatial resolution surfaces, it is still important to generate these datasets because without these surfaces we will not be able to determine the impact of climate change. Creating climate surfaces is important, however, we also need to understand the uncertainty when utilizing these in ecological niche modelling, since uncertainty can alter the outcome of a species potential habitat. Future research is still needed to find the optimal covariate for each variable, which will lead to producing a better climate surface. It is also very important to try to increase the temporal resolution (monthly to daily) since this can lead to applying it to more direct human applications such as in public health. In the summer of 2003, the increase in number of heat waves had a huge impact, such that the average mortality rate increased. Europe had more than 15,000 additional deaths caused by these heat waves, which makes it critical to understand the climate change/shift at a global scale. This topic is one of the biggest contributions that can be made. Another example being food security, there's a question of how and what crops should be grown due to

climate change. There are many routes that this dataset/product can make not only to the biodiversity community, but also as shown in many research areas.

REFERENCES

Allan, R. and T. Ansell (2006). "A new globally complete monthly historical gridded mean sea level pressure dataset (HadSLP2): 1850-2004." Journal of Climate **19**(22): 5816-5842.

Alvarez, O., Q. Guo, R. C. Klinger, W. Li and P. Doherty (2013). "Comparison of elevation and remote sensing derived products as auxiliary data for climate surface interpolation." International Journal of Climatology.

Austin, M. (2007). "Species distribution models and ecological theory: a critical assessment and some possible new approaches." Ecological modelling **200**(1): 1-19.

Beale, C. M. and J. J. Lennon (2012). Incorporating uncertainty in predictive species distribution modelling.

Berthier, E., Y. Arnaud, C. Vincent and F. Remy (2006). "Biases of SRTM in high-mountain areas: Implications for the monitoring of glacier volume changes." Geophysical Research Letters **33**(8).

Bonan, G. B., S. Levis, S. Sitch, M. Vertenstein and K. W. Oleson (2003). "A dynamic global vegetation model for use with climate models: concepts and description of simulated vegetation dynamics." Global Change Biology **9**(11): 1543-1566.

Bourgine, B. and N. Baghdadi (2005). "Assessment of C-band SRTM DEM in a dense equatorial forest zone." Comptes Rendus Geoscience **337**(14): 1225-1234.

Brooks, H. E. and C. A. Doswell (1993). "New technology and numerical weather prediction—a wasted opportunity?" Weather **48**(6): 173-177.

Busby, J. R. (1986). "A Biogeoclimatic Analysis of Nothofagus-Cunninghamii (Hook) Oerst in Southeastern Australia." Australian Journal of Ecology **11**(1): 1-7.

Carpenter, G., A. N. Gillison and J. Winter (1993). "Domain - a Flexible Modeling Procedure for Mapping Potential Distributions of Plants and Animals." Biodiversity and Conservation **2**(6): 667-680.

Chen, H., S. L. Guo, C. Y. Xu and V. P. Singh (2007). "Historical temporal trends of hydro-climatic variables and runoff response to climate variability and their relevance in water resource management in the Hanjiang basin." Journal of Hydrology **344**(3-4): 171-184.

Christensen, J. H., T. R. Carter and F. Giorgi (2002). "PRUDENCE employs new methods to assess European climate change." Eos, Transactions American Geophysical Union **83**(13): 147-147.

Daly, C. (2006). "Guidelines for assessing the suitability of spatial climate data sets." International Journal of Climatology **26**(6): 707-721.

Daly, C., W. P. Gibson, G. H. Taylor, G. L. Johnson and P. Pasteris (2002). "A knowledge-based approach to the statistical mapping of climate." Climate Research **22**(2): 99-113.

Daly, C., M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis and P. P. Pasteris (2008). "Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States." International Journal of Climatology **28**(15): 2031-2064.

Daly, C., G. H. Taylor, W. P. Gibson, T. W. Parzybok, G. L. Johnson and P. A. Pasteris (2000). "High-quality spatial climate data sets for the United States and beyond." Transactions of the Asae **43**(6): 1957-1962.

Damschen, E. I., S. Harrison and J. B. Grace (2010). "Climate change effects on an endemic-rich edaphic flora: resurveying Robert H. Whittaker's Siskiyou sites (Oregon, USA)." Ecology **91**(12): 3609-3619.

Dessai, S., M. Hulme, R. Lempert and R. Pielke Jr (2009). "Climate prediction: a limit to adaptation." Adapting to climate change: thresholds, values, governance: 64-78.
Diodato, N. (2005). "The influence of topographic co-variables on the spatial variability of precipitation over small regions of complex terrain." International Journal of Climatology **25**(3): 351-363.
Doherty, P., Q. Guo and O. Alvarez (2012). "Expert versus Machine: A Comparison of Two Suitability Models for Emergency Helicopter Landing Areas in Yosemite National Park." The Professional Geographer **65**(3): 466-481.
Duncan, D. B. (1955). "Multiple range and multiple F tests." Biometrics **11**(1): 1-42.
Elith, J., C. H. Graham, R. P. Anderson, M. Dudik, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M. Overton, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberon, S. Williams, M. S. Wisz and N. E. Zimmermann (2006). "Novel methods improve prediction of species' distributions from occurrence data." Ecography **29**(2): 129-151.
ESRI (2009). "ArcMap 9.3." ESRI, Redlands, California.
FAO (2001). "FAOCLIM 2.0 A World-Wide Agroclimatic Database " Food and Agriculture Organization of the United Nations.
FAO (2001). "FAOCLIM 2.0 A World-Wide Agroclimatic Database." Food and Agriculture Organization of the United Nations.
Farr, T. G., P. A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth, D. Seal, S. Shaffer, J. Shimada, J. Umland, M. Werner, M. Oskin, D. Burbank and D. Alsdorf (2007). "The shuttle radar topography mission." Reviews of Geophysics **45**(2).
Feria, T. P. and A. T. Peterson (2002). "Prediction of bird community composition based on point-occurrence data and inferential algorithms: a valuable tool in biodiversity assessments." Diversity and Distributions **8**(2): 49-56.
Fernandez, 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**(1).
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.
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): 1-15.
Fotheringham, A. S., M. E. Charlton and C. Brunsdon (1998). "Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis." Environment and Planning A **30**(11): 1905-1927.
Furrer, R., D. Nychka and S. Sain. (2011). "Package 'fields'." Retrieved May 1, 2012, from http://cran.rproject.org/web/packages/fields/index.html.
GBIF (2014). "Global Biodiversity Information Facility Data Portal."

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**(8): 1781-1793.
Gruber, A. and V. Levizzani (2008). "Assessment Global Precipitation Report, A Project of the Global Energy and Water Cycle Experiment (GEWEX) Radiation Panel GEWEX." 16.
Guisan, A., T. C. Edwards and T. Hastie (2002). "Generalized linear and generalized additive models in studies of species distributions: setting the scene." Ecological Modelling **157**(2-3): 89-100.
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 & Remote Sensing **76**(6): 701-712.
Guo, Q., Y. Liu and J. Wieczorek (2008). "Georeferencing locality descriptions and computing associated uncertainty using a probabilistic approach." International Journal of Geographical Information Science **22**(10): 1067-1090.
Guo, Q. H., M. Kelly and C. H. Graham (2005). "Support vector machines for predicting distribution of sudden oak death in California." Ecological Modelling **182**(1): 75-90.
Hancock, P. A. and M. F. Hutchinson (2006). "Automatic computation of hierarchical biquadratic smoothing splines with minimum GCV." Computers & Geosciences **32**(6): 834-845.
Hancock, P. A. and M. F. Hutchinson (2006). "Spatial interpolation of large climate data sets using bivariate thin plate smoothing splines." Environmental Modelling & Software **21**(12): 1684-1694.
Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones and M. New (2008). "A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006." Journal of Geophysical Research: Atmospheres **113**(D20): D20119.
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**(6): 751-777.
Heikkinen, R. K., M. Luoto, M. B. Araújo, R. Virkkala, W. Thuiller and M. T. Sykes (2006). "Methods and uncertainties in bioclimatic envelope modelling under climate change." Progress in Physical Geography **30**(6): 751-777.
Heuvelink, G. B., P. A. Burrough and A. Stein (1999). "Propagation of errors in spatial modelling with GIS." International Journal of Geographical Information System **3**(4): 303-322.
Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones and A. Jarvis (2005). "Very high resolution interpolated climate surfaces for global land areas." International Journal of Climatology **25**(15): 1965-1978.
Hofstra, N., M. Haylock, M. New, P. Jones and C. Frei (2008). "Comparison of six methods for the interpolation of daily, European climate data." Journal of Geophysical Research-Atmospheres **113**(D21): -.
Hong, Y., H. A. Nix, M. F. Hutchinson and T. H. Booth (2005). "Spatial interpolation of monthly mean climate data for China." International Journal of Climatology **25**(10): 1369-1379.

Honkavaara, E., R. Arbiol, L. Markelin, L. Martinez, M. Cramer, S. Bovet, L. Chandelier, R. Ilves, S. Klonus and P. Marshal (2009). "Digital airborne photogrammetry—A new tool for quantitative remote sensing?—A state-of-the-art review on radiometric aspects of digital photogrammetric images." Remote Sensing **1**(3): 577-605.
Hutchinson, M. F. (1995). "Interpolating Mean Rainfall Using Thin-Plate Smoothing Splines." International Journal of Geographical Information Systems **9**(4): 385-403.
Hutchinson, M. F. and P. E. Gessler (1994). "Splines - More Than Just a Smooth Interpolator." Geoderma **62**(1-3): 45-67.
Jarvis, A., H. Reuter, A. Nelson and E. Guevara (2008). "Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database. htt p." srtm. csi. cgiar. org.
Jeffrey, S. J., J. O. Carter, K. B. Moodie and A. R. Beswick (2001). "Using spatial interpolation to construct a comprehensive archive of Australian climate data." Environmental Modelling & Software **16**(4): 309-330.
Kalnay, E. and M. Cai (2003). "Impact of urbanization and land-use change on climate." Nature **423**(6939): 528-531.
Kearney, M. R., A. Shamakhy, R. Tingley, D. J. Karoly, A. A. Hoffmann, P. R. Briggs and W. P. Porter (2014). "Microclimate modelling at macro scales: a test of a general microclimate model integrated with gridded continental-scale soil and weather data." Methods in Ecology and Evolution: n/a-n/a.
Kohavi, R. (1995). "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection." INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE **14**: 1137-1145.
Kriticos, D. J. and A. Leriche (2010). "The effects of climate data precision on fitting and projecting species niche models." Ecography **33**(1): 115-127.
Leemans, R. and W. P. Cramer (1991). "The IIASA database for mean monthly values of temperature, precipitation, and cloudiness on a global terrestrial grid." INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS, Laxenburg, Austria.
Lenoir, J., J. C. Gegout, 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**(2): 295-303.
Li, W. and Q. Guo (2010). "A maximum entropy approach to one-class classification of remote sensing imagery." International Journal of Remote Sensing **31**(8): 2227-2235.
Li, W., Q. Guo and C. Elkan (2011). "Can we model the probability of presence of species without absence data?" Ecography **34**: 1-10.
Li, W. K. and Q. H. Guo (2010). "A maximum entropy approach to one-class classification of remote sensing imagery." International Journal of Remote Sensing **31**(8): 2227-2235.
Li, W. K., Q. H. Guo and C. Elkan (2011). "A Positive and Unlabeled Learning Algorithm for One-Class Classification of Remote-Sensing Data." Ieee Transactions on Geoscience and Remote Sensing **49**(2): 717-725.
Loarie, S. R., P. B. Duffy, H. Hamilton, G. P. Asner, C. B. Field and D. D. Ackerly (2009). "The velocity of climate change." Nature **462**(7276): 1052-U1111.
Lucas, R. M., N. Cronin, A. Lee, M. Moghaddam, C. Witte and P. Tickle (2006). "Empirical relationships between AIRSAR backscatter and LiDAR-derived forest biomass, Queensland, Australia." Remote Sensing of Environment **100**(3): 407-425.

Mandelik, Y., U. Roll and A. Fleischer (2010). "Cost-efficiency of biodiversity indicators for Mediterranean ecosystems and the effects of socio-economic factors." Journal of Applied Ecology **47**(6): 1179-1188.
Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier and B. Nijssen (2002). "A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States." Journal of Climate **15**(22): 3237-3251.
Mbogga, M. S., A. Hamann and T. L. Wang (2009). "Historical and projected climate data for natural resource management in western Canada." Agricultural and Forest Meteorology **149**(5): 881-890.
McNemar, Q. (1947). "Note on the sampling error of the difference between correlated proportions or percentages." Psychometrika **12**(2): 153-157.
McPherson, G. R. and J. F. Weltzin (2000). "Disturbance and climate change in United States/Mexico borderland plant communities: a state-of-the-knowledge review."
Misra, P. and P. Enge (2006). Global Positioning System: Signals, Measurements and Performance Second Edition, Lincoln, MA: Ganga-Jamuna Press.
Morin, X. and W. Thuiller (2009). "Comparing niche- and process-based models to reduce prediction uncertainty in species range shifts under climate change." Ecology **90**(5): 1301-1313.
New, M., M. Hulme and P. Jones (1999). "Representing twentieth-century space-time climate variability. Part I: Development of a 1961-90 mean monthly terrestrial climatology." Journal of Climate **12**(3): 829-856.
New, M., D. Lister, M. Hulme and I. Makin (2002). "A high-resolution data set of surface climate over global land areas." Climate Research **21**(1): 1-25.
Newbold, T. (2010). "Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models." Progress in Physical Geography **34**(1): 3-22.
NOAA. (2012). "Downloading Gridded Rainfall Data." Retrieved May 1, 2012, from http://water.weather.gov/precip/download.php.
Olson, D. M., E. Dinerstein, E. D. Wikramanayake, N. D. Burgess, G. V. Powell, E. C. Underwood, J. A. D'amico, I. Itoua, H. E. Strand and J. C. Morrison (2001). "Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity." BioScience **51**(11): 933-938.
Pearson, R. G., T. P. Dawson and C. Liu (2004). "Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data." Ecography **27**(3): 285-298.
Pearson, R. G., C. J. Raxworthy, M. Nakamura and A. Townsend Peterson (2007). "Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar." Journal of Biogeography **34**(1): 102-117.
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**(1): 135-144.
Peterson, T. C. and R. S. Vose (1997). "An overview of the global historical climatology network temperature database." Bulletin of the American Meteorological Society **78**(12): 2837-2849.

Phillips, S. J., R. P. Anderson and R. E. Schapire (2006). "Maximum entropy modeling of species geographic distributions." Ecological Modelling **190**(3-4): 231-259.
Phillips, S. J., M. Dudik, J. Elith, C. H. Graham, A. Lehmann, J. Leathwick and S. Ferrier (2009). "Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data." Ecological Applications **19**(1): 181-197.
Price, D. T., D. W. McKenney, I. A. Nalder, M. F. Hutchinson and J. L. Kesteven (2000). "A comparison of two statistical methods for spatial interpolation of Canadian monthly mean climate data." Agricultural and Forest Meteorology **101**(2–3): 81-94.
Purcell, K. L., A. K. Mazzoni, S. R. Mori and B. B. Boroski (2009). "Resting structures and resting habitat of fishers in the southern Sierra Nevada, California." Forest Ecology and Management **258**(12): 2696-2706.
Pyke, G. H. and P. R. Ehrlich (2010). "Biological collections and ecological/environmental research: a review, some observations and a look to the future." Biological Reviews **85**(2): 247-266.
Richardson, J. J., L. M. Moskal and S.-H. Kim (2009). "Modeling approaches to estimate effective leaf area index from aerial discrete-return LIDAR." Agricultural and Forest Meteorology **149**(6): 1152-1160.
Robine, J.-M., S. L. K. Cheung, S. Le Roy, H. Van Oyen, C. Griffiths, J.-P. Michel and F. R. Herrmann (2008). "Death toll exceeded 70,000 in Europe during the summer of 2003." Comptes Rendus Biologies **331**(2): 171-178.
Rodriguez, E., C. S. Morris and J. E. Belz (2006). "A global assessment of the SRTM performance." Photogrammetric Engineering & Remote Sensing **72**(3): 249-260.
Rodriguez, E., C. S. Morris and J. E. Belz (2006). "A global assessment of the SRTM performance." Photogrammetric Engineering and Remote Sensing **72**(3): 249-260.
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**(3): 243-251.
Ruggiero, L. F., K. B. Aubry, S. W. Buskirk, J. Lyon and W. J. Zielinski (1994). "The scientific basis for conserving forest carnivores: American marten, fisher, lynx and wolverine in the western United States." General Technical Report US Department of Agriculture Forest Service: Pagination missing-please provide.
Ruggiero, L. F., K. B. Aubry, S. W. Buskirk, L. J. Lyon and W. J. Zielinski (1994). "The scientific basis for conserving forest carnivores: American marten, fisher, lynx, and wolverine in the western United States."
Sexton, J. P., P. J. McIntyre, A. L. Angert and K. J. Rice (2009). "Evolution and ecology of species range limits."
Sharples, J. J., M. F. Hutchinson and D. R. Jellett (2005). "On the Horizontal Scale of Elevation Dependence of Australian Monthly Precipitation." Journal of Applied Meteorology **44**(12): 1850-1865.
Sheffield, J., A. D. Ziegler, E. F. Wood and Y. B. Chen (2004). "Correction of the high-latitude rain day anomaly in the NCEP-NCAR reanalysis for land surface hydrological modeling." Journal of Climate **17**(19): 3814-3828.
Stahl, K., R. D. Moore, J. A. Floyer, M. G. Asplin and I. G. McKendry (2006). "Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density." Agricultural and Forest Meteorology **139**(3-4): 224-236.

Stockwell, D. and D. Peters (1999). "The GARP modelling system: problems and solutions to automated spatial prediction." International Journal of Geographical Information Science **13**(2): 143-158.
Su, Y. and Q. Guo (2014). "A practical method for SRTM DEM correction over vegetated mountain areas." ISPRS Journal of Photogrammetry and Remote Sensing **87**: 216-228.
Sun, Q., C. Miao, Q. Duan, D. Kong, A. Ye, Z. Di and W. Gong (2014). "Would the 'real'observed dataset stand up? A critical examination of eight observed gridded climate datasets for China." Environmental Research Letters **9**(1): 015001.
Tait, A., R. Henderson, R. Turner and X. Zheng (2006). "Thin plate smoothing spline interpolation of daily rainfall for New Zealand using a climatological rainfall surface." International Journal of Climatology **26**(14): 2097-2115.
Thornton, P. E., S. W. Running and M. A. White (1997). "Generating surfaces of daily meteorological variables over large regions of complex terrain." Journal of Hydrology **190**(3-4): 214-251.
Thornton, P. K., P. G. Jones, G. Alagarswamy and J. Andresen (2009). "Spatial variation of crop yield response to climate change in East Africa." Global Environmental Change **19**(1): 54-65.
Trabucco, A., R. J. Zomer, D. A. Bossio, O. van Straaten and L. V. Verchot (2008). "Climate change mitigation through afforestation/reforestation: A global analysis of hydrologic impacts with four case studies." Agriculture Ecosystems & Environment **126**(1-2): 81-97.
Vörösmarty, C., M. Routhier, A. Wright, T. Baker, C. Fernandez-Jauregui and M. Donoso (1998). "A regional hydrometeorological data network for South America, Central America, and the Caribbean (R-HydroNET V1. 0)." Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Durham, NH.
Wahba, G. (1990). Spline models for observational data. Philadelphia, Pa., Society for Industrial and Applied Mathematics.
Wahba, G. and J. Wendelberger (1980). "Some New Mathematical-Methods for Variational Objective Analysis Using Splines and Cross Validation." Monthly Weather Review **108**(8): 1122-1143.
Walther, G. R., E. Post, P. Convey, A. Menzel, C. Parmesan, T. J. C. Beebee, J. M. Fromentin, O. Hoegh-Guldberg and F. Bairlein (2002). "Ecological responses to recent climate change." Nature **416**(6879): 389-395.
Wang, T., A. Hamann, D. L. Spittlehouse and T. Q. Murdock (2011). "ClimateWNA—High-Resolution Spatial Climate Data for Western North America." Journal of Applied Meteorology and Climatology **51**(1): 16-29.
Weydahl, D., J. Sagstuen, Ø. Dick and H. Rønning (2007). "SRTM DEM accuracy assessment over vegetated areas in Norway." International Journal of Remote Sensing **28**(16): 3513-3527.
Wieczorek, J., Q. Guo and R. Hijmans (2004). "The point-radius method for georeferencing locality descriptions and calculating associated uncertainty." International Journal of Geographical Information Science **18**(8): 745-767.
Wing, M. G., A. Eklund and L. D. Kellogg (2005). "Consumer-grade global positioning system (GPS) accuracy and reliability." Journal of Forestry **103**(4): 169-173.
Yates, F. (1934). "Contingency tables involving small numbers and the χ 2 test." Supplement to the Journal of the Royal Statistical Society: 217-235.

You, Q. L., S. C. Kang, N. Pepin and Y. P. Yan (2008). "Relationship between trends in temperature extremes and elevation in the eastern and central Tibetan Plateau, 1961-2005." Geophysical Research Letters **35**(4): -.
Zhao, F., R. A. Sweitzer, Q. Guo and M. Kelly (2012). "Characterizing habitats associated with fisher den structures in the Southern Sierra Nevada, California using discrete return lidar." Forest Ecology and Management **280**(0): 112-119.
Zielinski, W. J., J. R. Dunk, J. S. Yaeger and D. W. LaPlante (2010). "Developing and testing a landscape-scale habitat suitability model for fisher (< i> Martes pennanti</i>) in forests of interior northern California." Forest ecology and management **260**(9): 1579-1591.
Zielinski, W. J., R. L. Truex, F. V. Schlexer, L. A. Campbell and C. Carroll (2005). "Historical and contemporary distributions of carnivores in forests of the Sierra Nevada , California , USA." Journal of Biogeography **32**(8): 1385-1407.
Zielinski, W. J., R. L. Truex, F. V. Schlexer, L. A. Campbell and C. Carroll (2005). "Historical and contemporary distributions of carnivores in forests of the Sierra Nevada, California, USA." Journal of Biogeography **32**(8): 1385-1407.

APPENDIX A

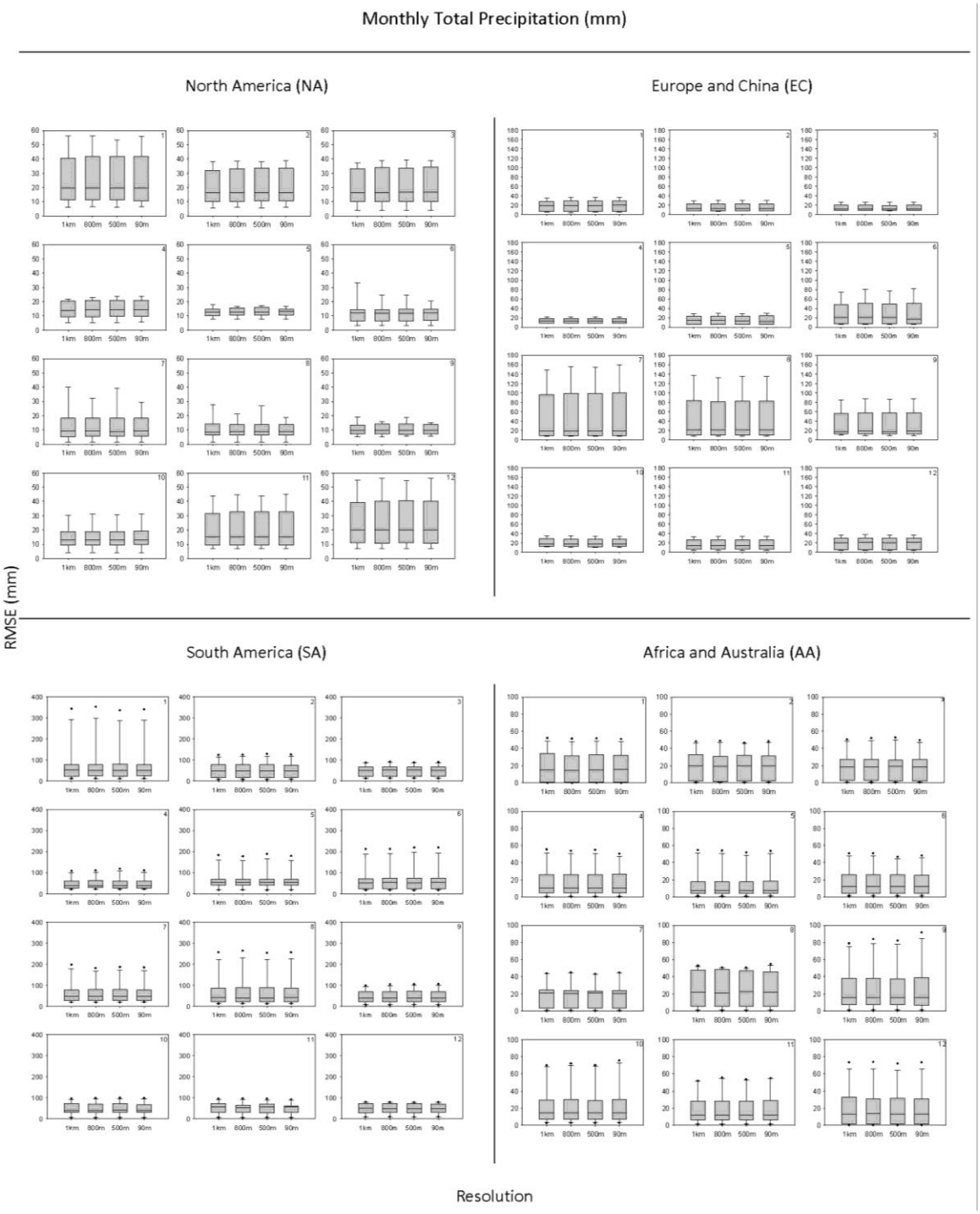


Figure 1: Monthly total precipitation RMSE (mm). Each quadrant represents a different continental group, with each graph representing a different month.

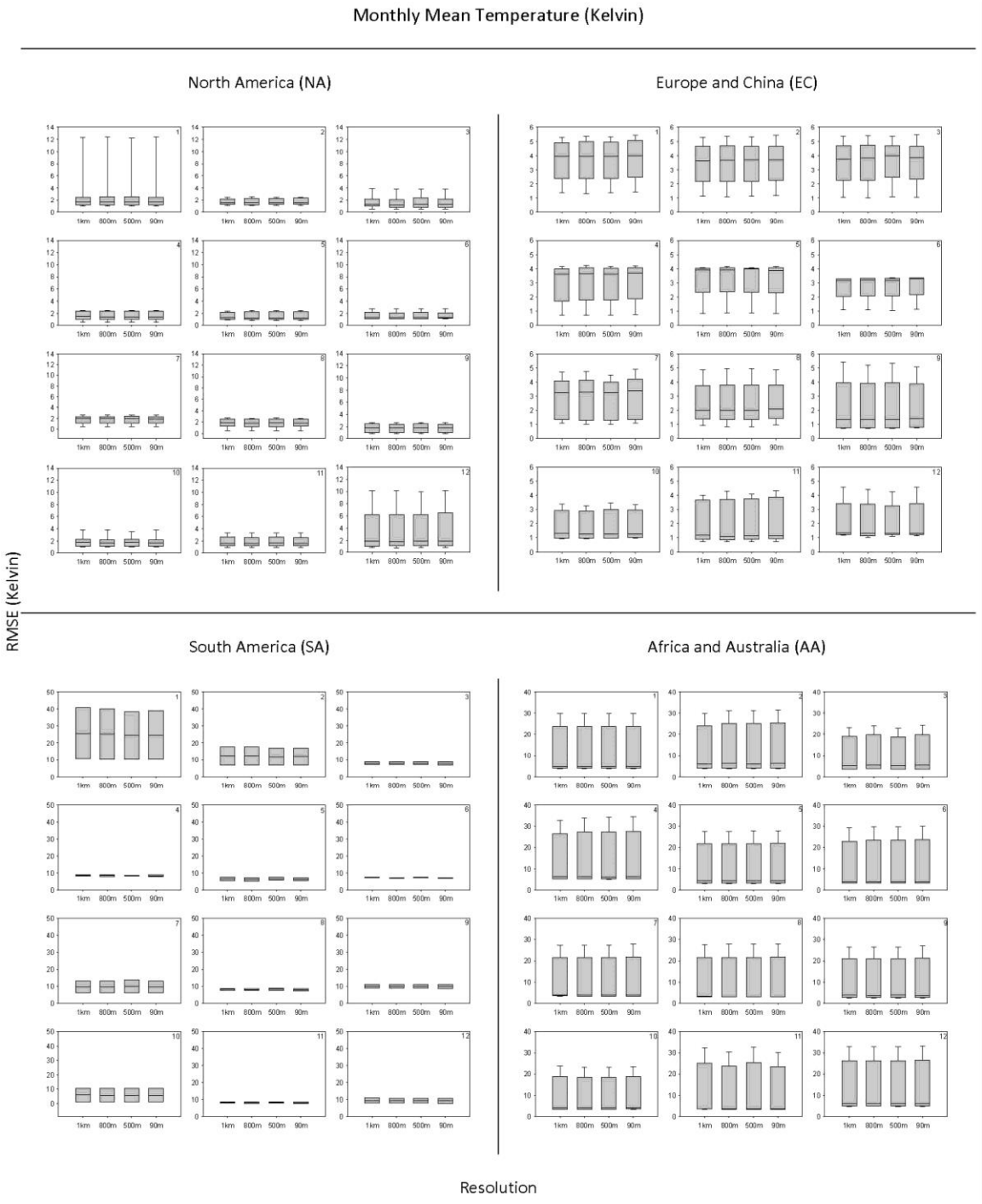


Figure 2:Monthly mean temperature RMSE (K). Each quadrant represents a different continental group, with each graph representing a different month.

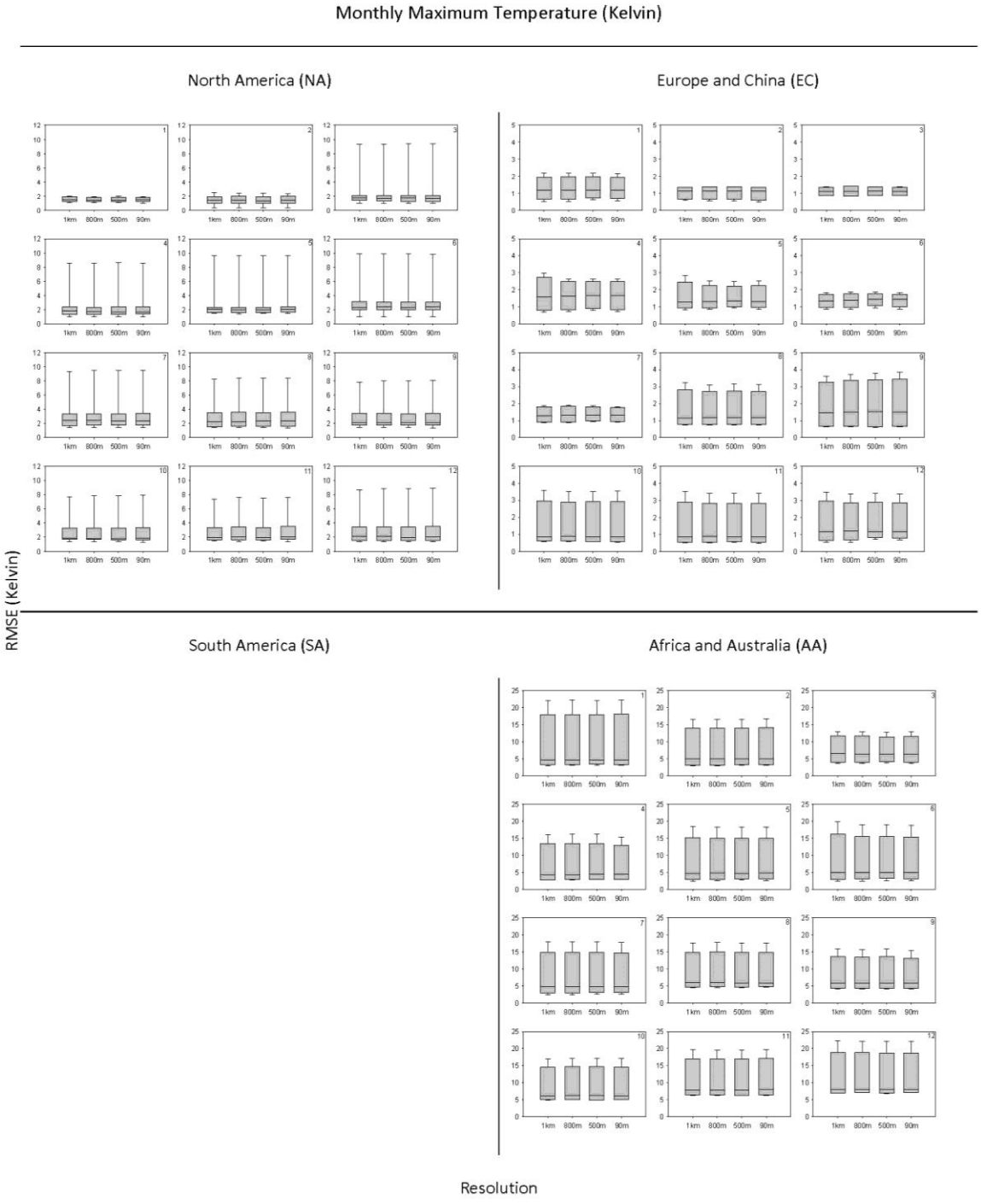


Figure 3:Monthly maximum temperature RMSE (K). Each quadrant represents a different continental group, with each graph representing a different month.

Figure 4: Monthly minimum temperature RMSE (K). Each quadrant represents a different continental group, with each graph representing a different month.

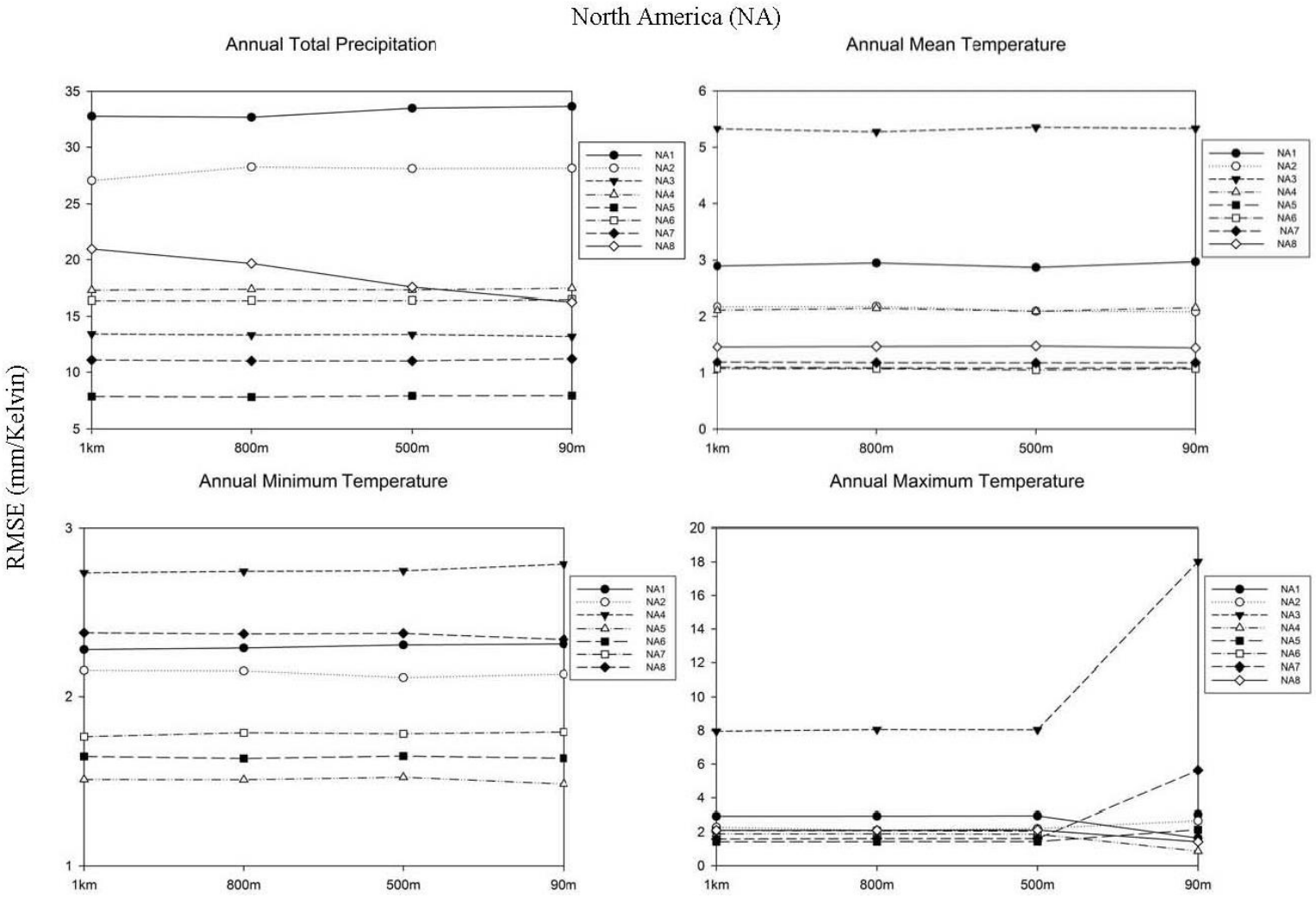


Figure 5: Total annual RMSE for **Co-Kriging** (for precipitation units are in mm and for temperature they are in K); each quadrant shows the available variables and display search study area in finer detail.

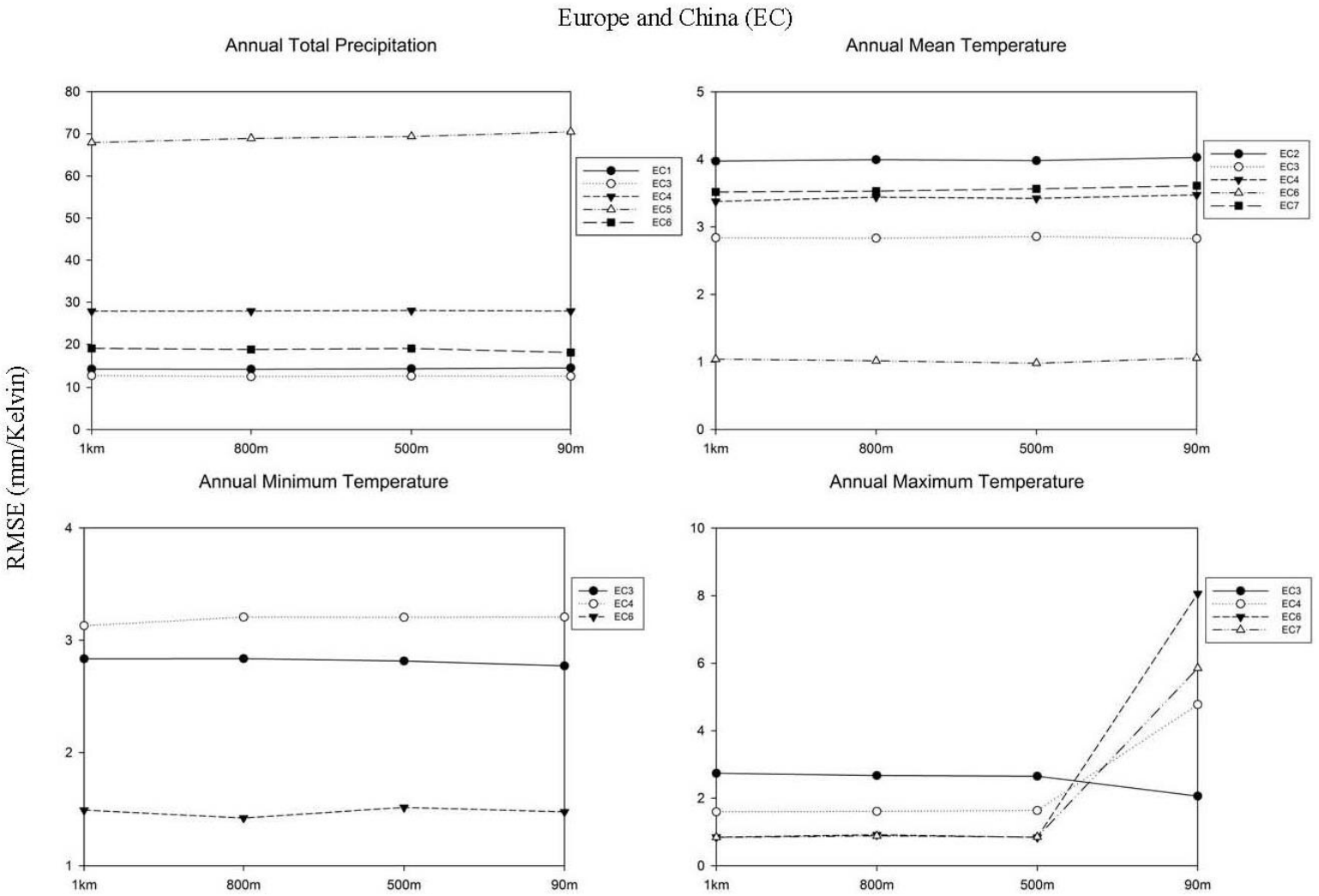


Figure 5: (Cont. Quadrant II):

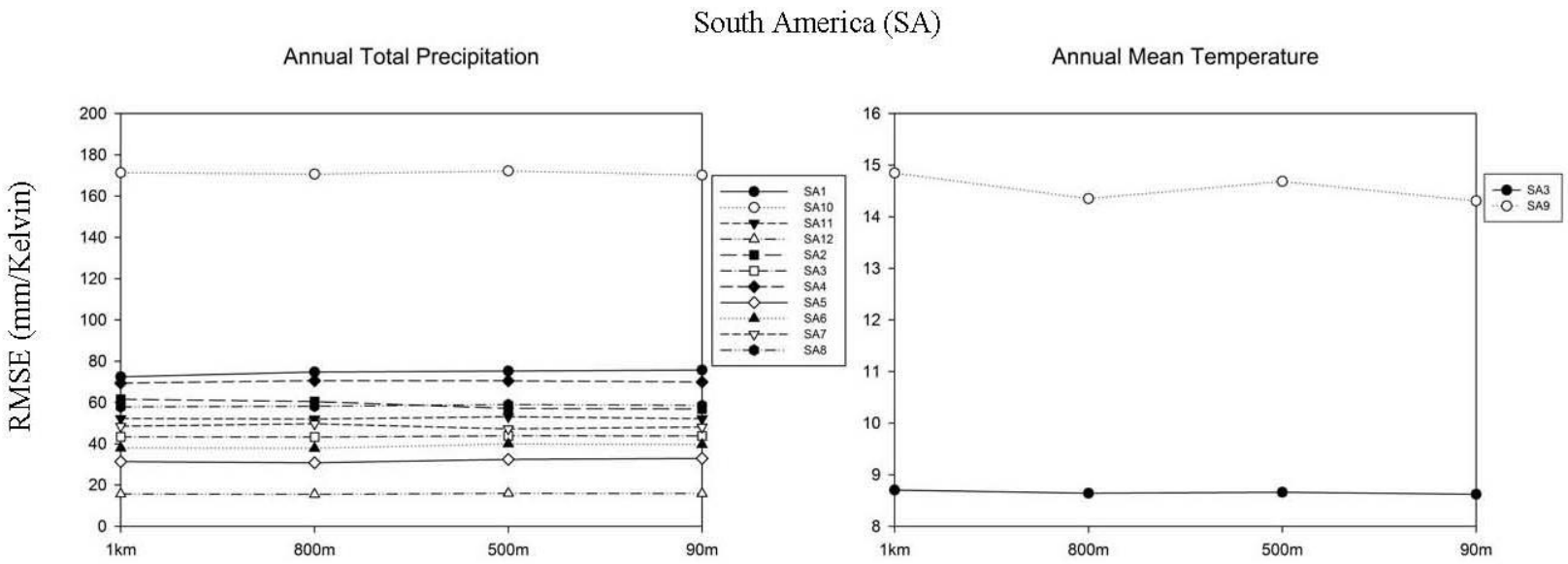


Figure 5: (Cont. Quadrant III):

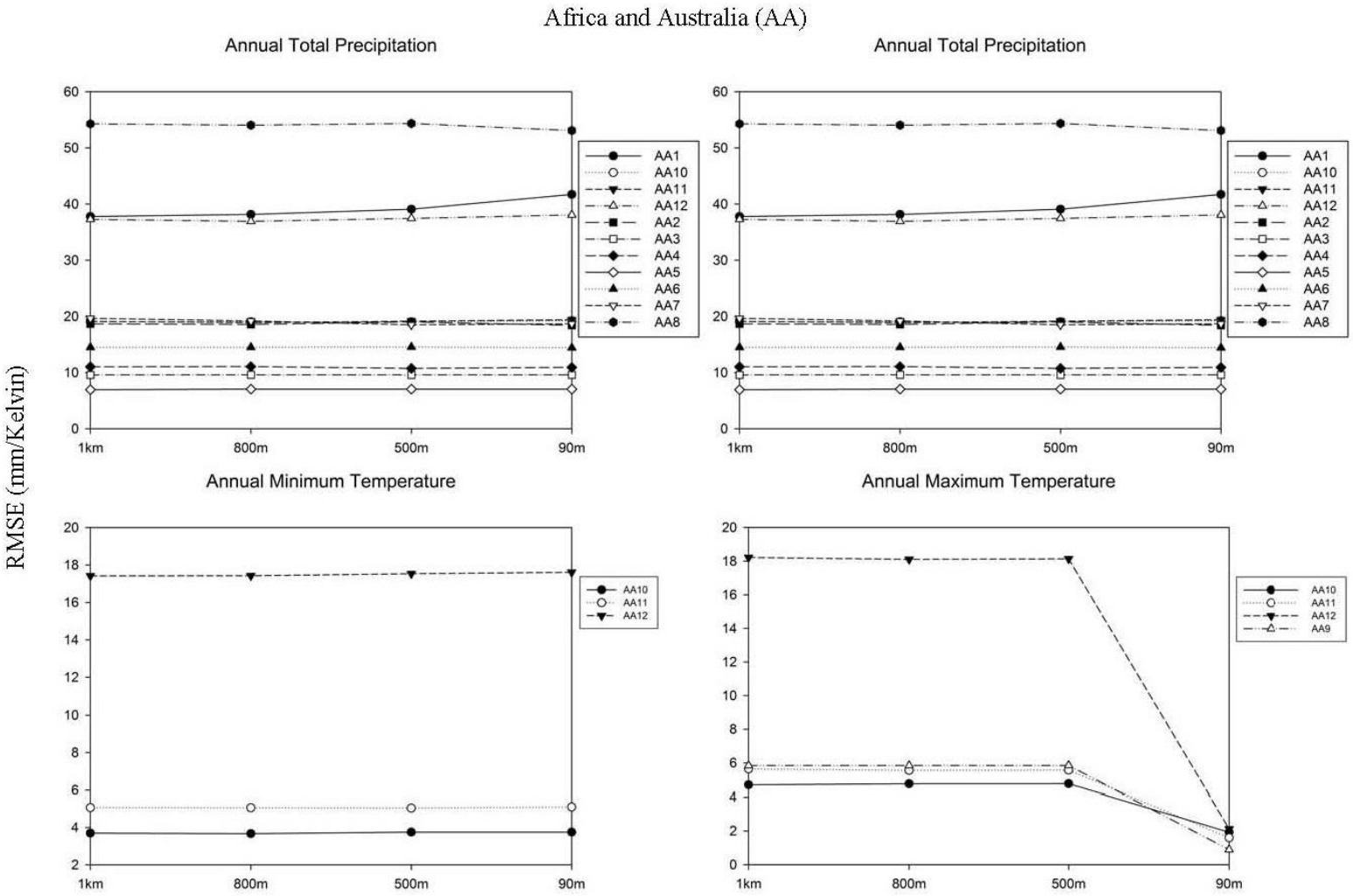


Figure 5: (Cont. Quadrant IV):

APPENDIX B

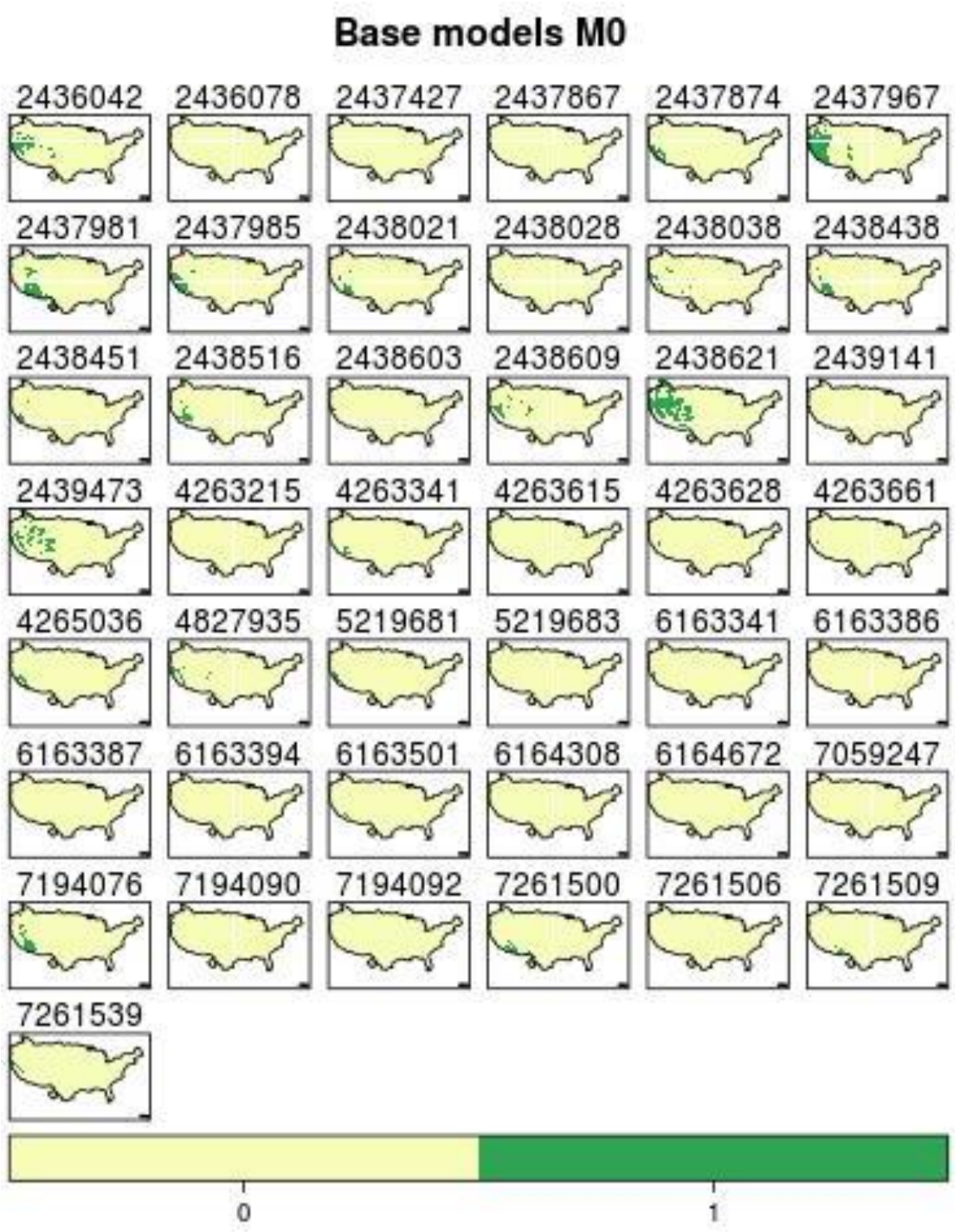


Figure 1: base models which was converted to binary.

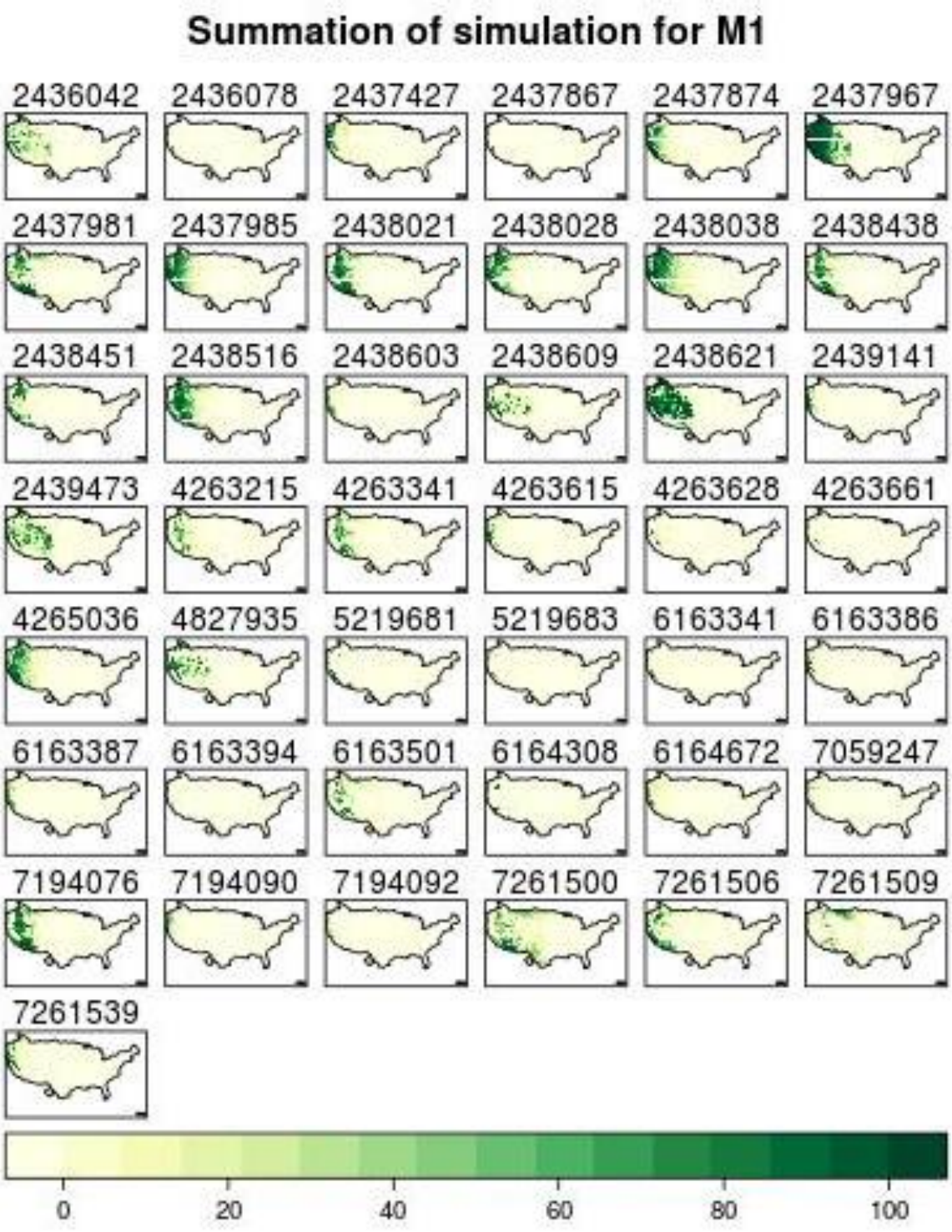


Figure 2: All the simulation models, which was converted to binary than sum for uncertainty on climate variables.

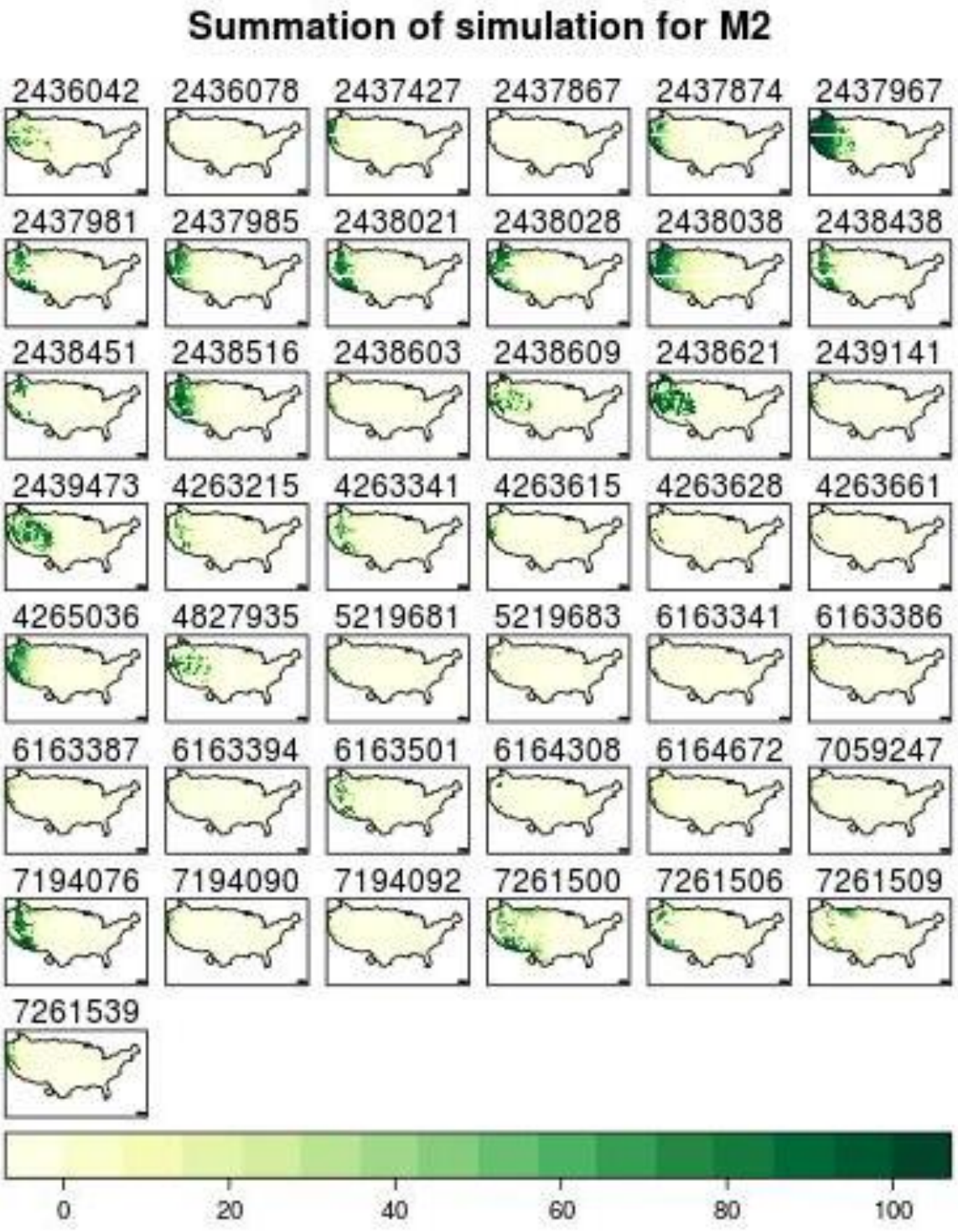


Figure 3: All the simulation models, which was converted to binary than sum for uncertainty on climate variables and DEM.

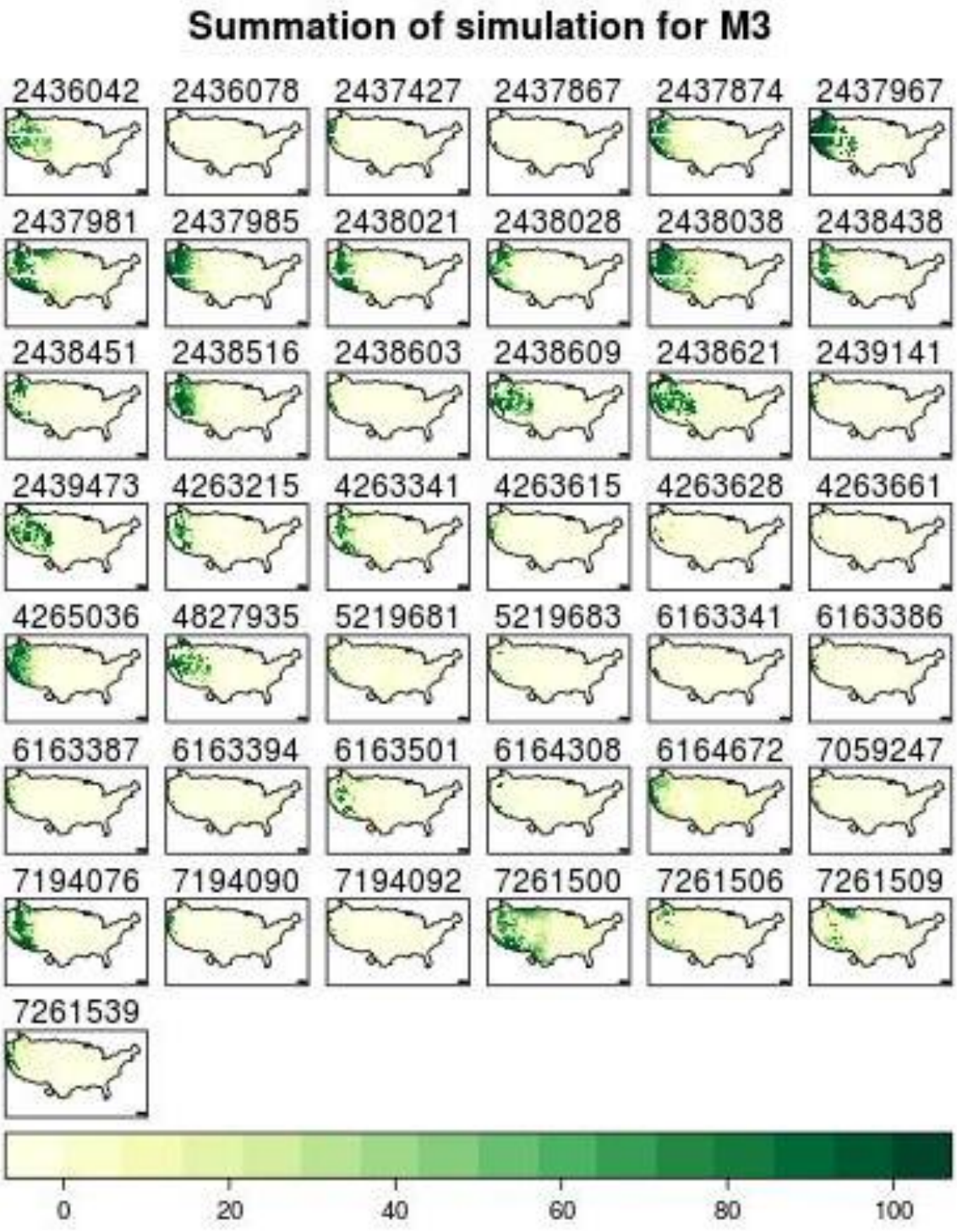


Figure 4: All the simulation models which was converted to binary than sum for uncertainty on climate variables and point locality.

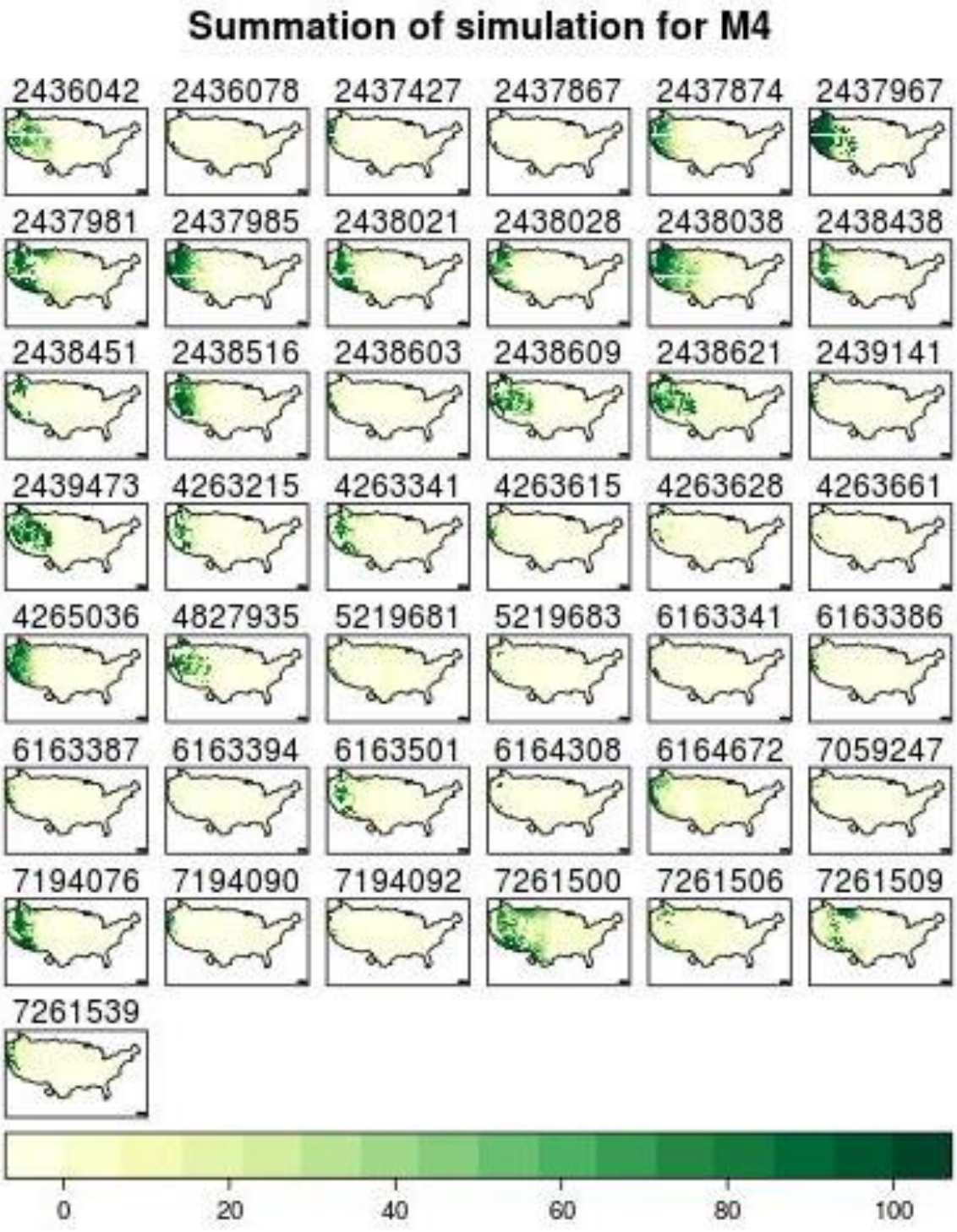


Figure 5: All the simulation models which was converted to binary than sum for uncertainty on all (climate variables, DEM, and point locality).

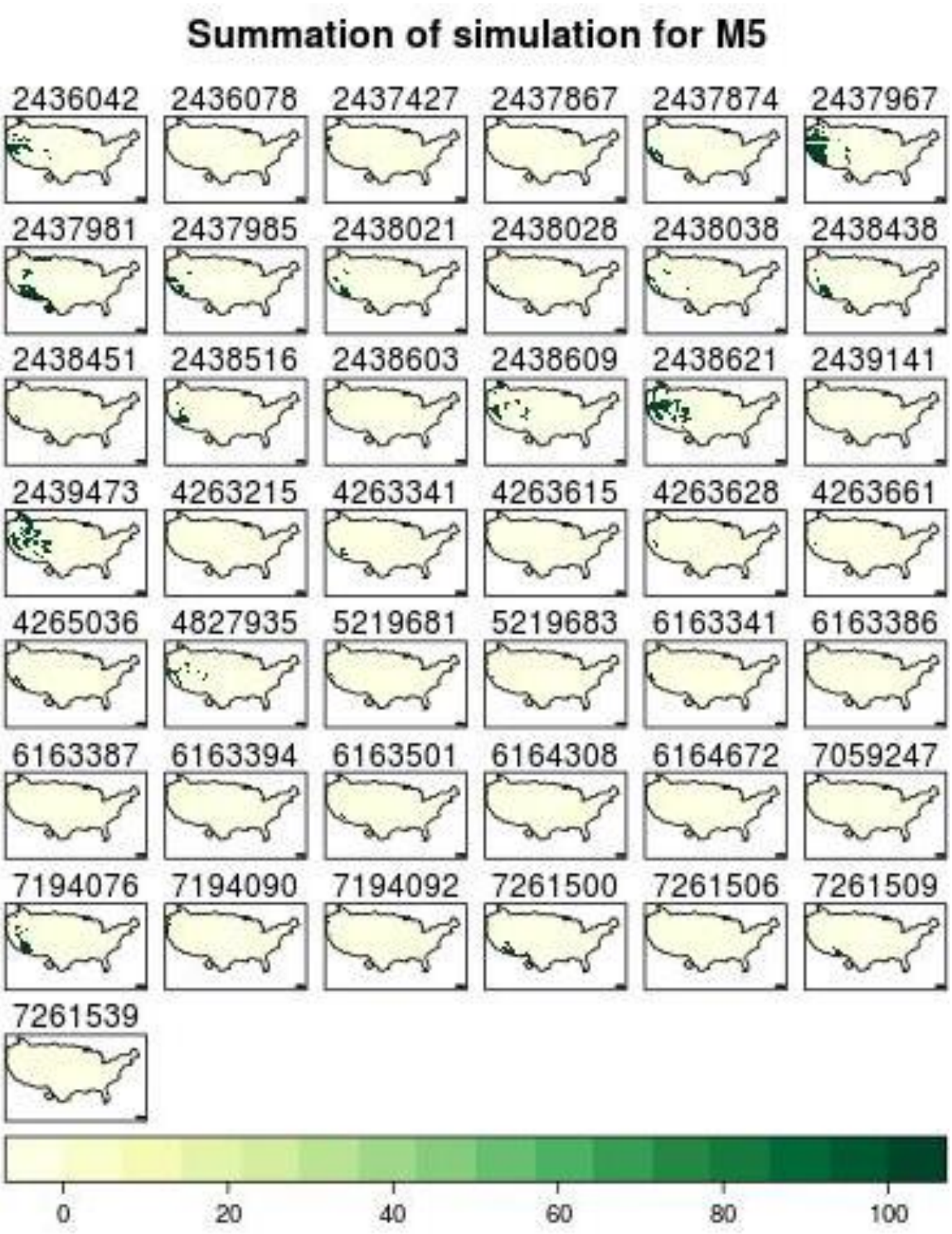


Figure 6: All the simulation models which was converted to binary than sum for uncertainty on DEM.

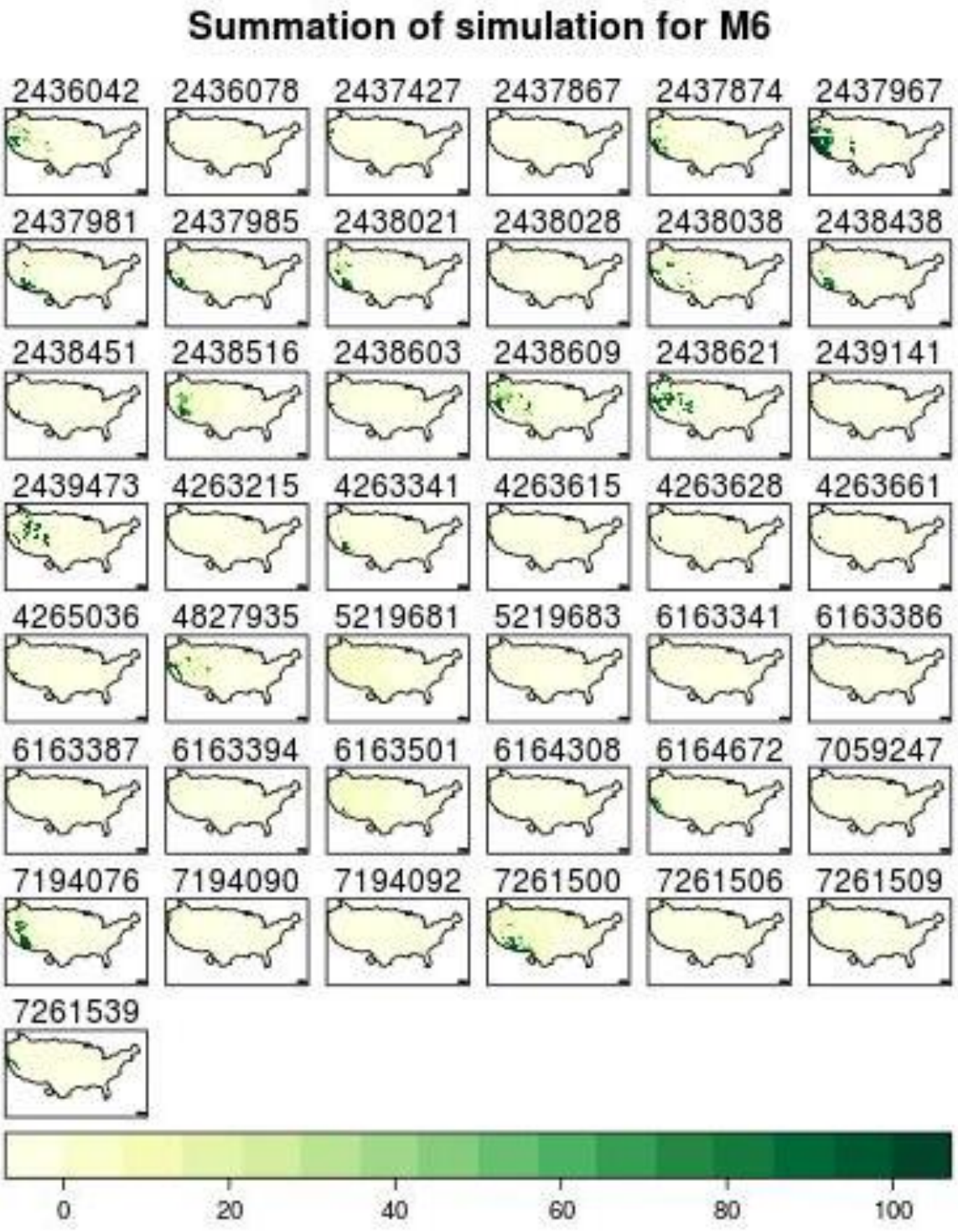


Figure 7: All the simulation models which was converted to binary than sum for uncertainty on DEM and point locality.

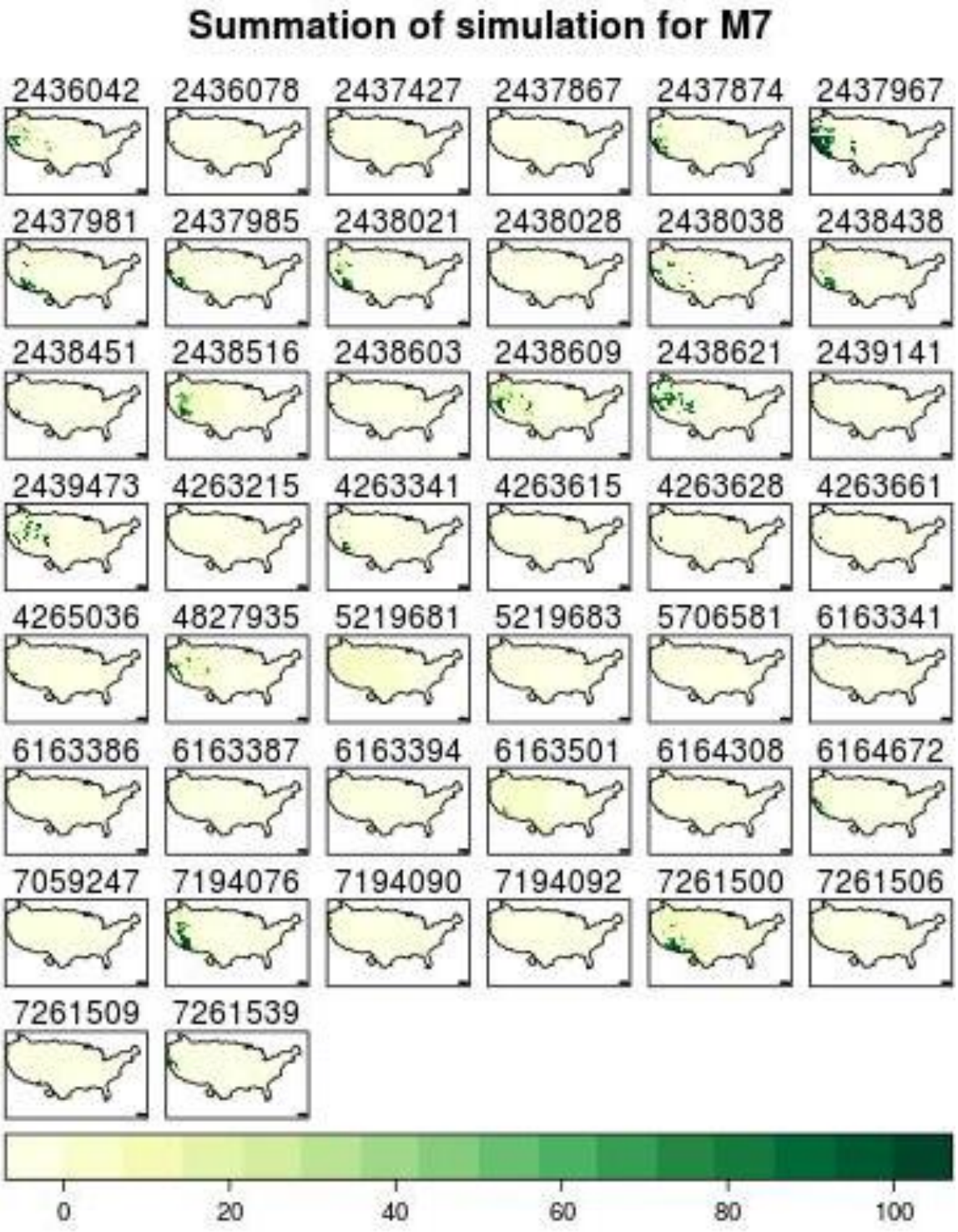


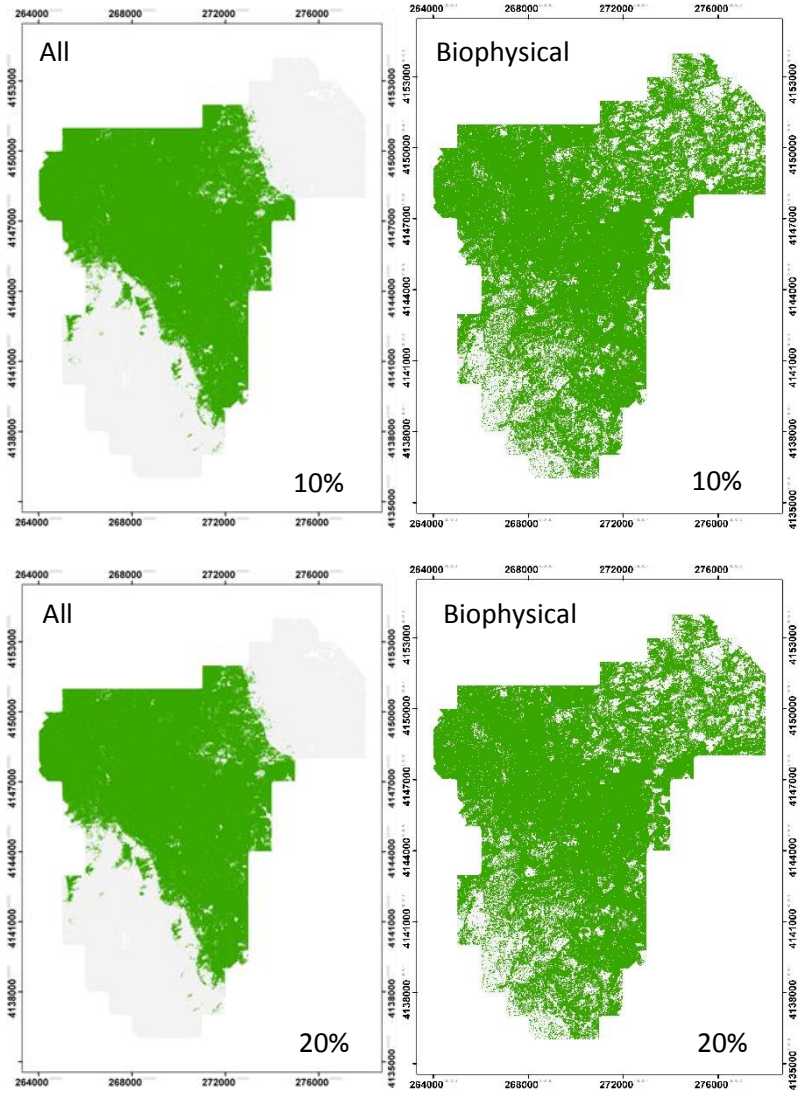
Figure 8: All the simulation models which was converted to binary than sum for uncertainty on point locality.

Table 1: All 43 species ID, scientific name, category, group, and the number of unique localities.

Species ID	Species	Category	Group	Count
2436042	Sorex vagrans Baird, 1857	Shrew	Broad	172
2436078	Sorex ornatus Merriam, 1895	Shrew	Broad	107
2437427	Tamias siskiyou (A. H. Howell, 1922)	Chipmunk	Narrow	168
2437867	Reithrodontomys raviventris Dixon, 1908	Mouse	Narrow	372
2437874	Reithrodontomys megalotis (Baird, 1858)	Mouse	Broad	432
2437967	Peromyscus maniculatus (Wagner, 1845)	Mouse	Broad	2006
2437981	Peromyscus eremicus (Baird, 1858)	Mouse	Broad	257
2437985	Peromyscus truei (Shufeldt, 1885)	Mouse	Broad	580
2438021	Peromyscus crinitus (Merriam, 1891)	Mouse	Broad	189
2438028	Peromyscus californicus (Gambel, 1848)	Mouse	Moderate	511
2438038	Peromyscus boylii (Baird, 1855)	Mouse	Broad	682
2438438	Neotoma lepida Thomas, 1893	Woodrat	Broad	251
2438451	Neotoma bryanti Merriam, 1887	Woodrat	Narrow	214
2438516	Onychomys torridus (Coues, 1874)	Woodrat	Broad	155
2438603	Microtus californicus (Peale, 1848)	Vole	Moderate	1665
2438609	Microtus montanus (Peale, 1848)	Vole	Broad	131
2438621	Microtus longicaudus (Merriam, 1888)	Vole	Broad	217
2439141	Clethrionomys californicus (Merriam, 1890)	Vole	Moderate	422
2439473	Zapus princeps J. A. Allen, 1893	Mouse	Broad	125
4263215	Dipodomys panamintinus subsp. mohavensis Grinnell, 1918	Kangaroo Rat	Narrow	182
4263341	Chaetodipus formosus subsp. mohavensis Huey, 1938	Pocket mouse	Broad	185
4263615	Tamias sonomae subsp. sonomae Grinnell, 1915	Chipmunk	Moderate	218
4263628	Tamias speciosus subsp. frater J. A. Allen, 1890	Chipmunk	Broad	182
4263661	Tamias amoenus subsp. monoensis Grinnell & Storer, 1916	Chipmunk	Broad	394
4265036	Neotoma macrotis Thomas, 1893	Woodrat	Broad	336
4827935	Callospermophilus lateralis Merriam, 1897 (?)	Squirrels	Broad	296
5219681	Sciurus carolinensis Gmelin, 1788	Squirrels	Broad	210
5219683	Sciurus niger Linnaeus, 1758	Squirrels	Broad	171
6163341	Sorex ornatus salicornicus von Bloeker, 1932	Shrew	Broad	224
6163386	Sorex trowbridgii humboldtensis Jackson, 1922	Shrew	Broad	486
6163387	Sorex trowbridgii montereyensis Merriam, 1895	Shrew	Broad	310
6163394	Sorex vagrans halicoetes Grinnell, 1913	Shrew	Broad	168
6163501	Perognathus longimembris longimembris (Coues, 1875)	Pocket mouse	Broad	161
6164308	Vulpes macrotis mutica Merriam, 1902	Fox	Broad	546
6164672	Lynx rufus californicus Mearns, 1897	Bobcat	Broad	100
7059247	Microtus californicus californicus (Peale, 1848)	Vole	Moderate	582

7194076	Dipodomys merriami subsp. merriami	Kangaroo Rat	Broad	270
7194090	Sorex trowbridgii subsp. trowbridgii	Shrew	Moderate	155
7194092	Sorex vagrans subsp. vagrans	Shrew	Moderate	1786
7261500	Chaetodipus intermedius subsp. intermedius	Pocket mouse	Broad	117
7261506	Chaetodipus fallax subsp. fallax	Pocket mouse	Narrow	125
7261509	Chaetodipus baileyi subsp. baileyi	Pocket mouse	Broad	873
7261539	Tamias senex subsp. senex	Chipmunk	Broad	132

APPENDIX C



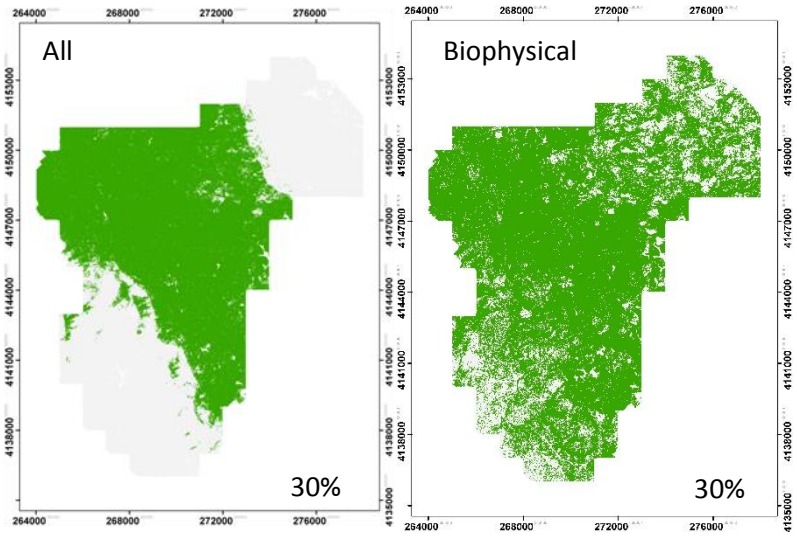


Figure 1: Model results after projection to a percentage of forest thinning.