THESIS

USING SPATIOTEMPORAL CORRELATIVE NICHE MODELS FOR EVALUATING THE EFFECTS OF CLIMATE CHANGE ON MOUNTAIN PINE BEETLE

Submitted by

Aaron M. Sidder

Graduate Degree Program in Ecology

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2015

Master's Committee:

Advisor: Melinda Laituri
Co-Advisor: Sunil Kumar

Jason Sibold

ABSTRACT

USING SPATIOTEMPORAL CORRELATIVE NICHE MODELS FOR EVALUATING THE EFFECTS OF CLIMATE CHANGE ON MOUNTAIN PINE BEETLE

Over the last decade western North America has experienced the largest mountain pine beetle (*Dendroctonus ponderosae* Hopkins) outbreak in recorded history and Rocky Mountain forests have been severely impacted. Although bark beetles are indigenous to North American forests, climate change has facilitated the beetle's expansion into previously unsuitable habitats. I used three correlative niche models (MaxEnt, Boosted Regression Trees, and Generalized Linear Models) to estimate: (i) the current potential distribution of the beetle in the U.S. Rocky Mountain region, (ii) how this extent has changed since historical outbreaks in the 1960s and 1970s, and (iii) how the potential distribution may be expected to change under future climate scenarios. Additionally, I evaluated the temporal transferability of the niche models by forecasting historical models and testing the model predictions using temporally independent outbreak data from the current outbreak. My results indicated that there has been a significant expansion of climatically suitable habitat over the past 50 years and that much of this expansion corresponds with an upward shift in elevation across the study area. Furthermore, my models indicate that drought was a more prominent driver of current outbreak than temperature, which suggests a change in the climatic signature between historical and current outbreaks. The current climatic niche of the mountain pine beetle includes increased precipitation, colder winter temperatures, and a later spring than the historical climatic niche, which reflects a shift into higher elevation habitats. Projections under future conditions suggest that there will be a large reduction in climatically suitable habitat for the beetle and that high-elevation forests will continue to become more susceptible to outbreak. While all three models generated reasonable predictions (AUC = 0.85 - 0.87), the generalized linear model correctly predicted a higher percentage of current outbreak localities when trained on historical data. My findings suggest that projects aiming to reduce omission error in estimates of future species responses may have greater predictive success with simpler, generalized models.

ACKNOWLEDGEMENTS

There are many people who helped make my graduate research possible and contributed to my success. First and foremost, I would like to thank my committee. Dr. Sunil Kumar has been a wonderful mentor and I am grateful for his patience and direction throughout this project and the course of my studies. Dr. Melinda Laituri has provided wisdom and support over the last few years and I thank her for guiding and challenging me to constantly improve my work; your suggestion to “walk a lot” has changed how I approach problems. Thanks to Dr. Jason Sibold for believing in me and giving me the opportunity to grow as an ecologist and spend a summer in some beautiful locales. I am deeply appreciative of the support from all of you.

Thanks to the Natural Resource Ecology Laboratory, the Warner College of Natural Resources (Roger and Connie Hoffer geospatial science scholarship), and the Colorado Mountain Club Foundation (Neil B. Kindig Fellowship) for funding my graduate studies. This support has meant a great deal during my time at Colorado State University (CSU).

Thanks to Tim Assal (U.S. Geological Survey, Graduate Degree Program in Ecology, CSU), Brian Howell and Justin Backsen (Forest Health Protection, U.S. Forest Service [USFS], Rocky Mountain Region), and Dick Halsey (Forest Health Protection, USFS, Boise Field Office) for their assistance in procuring historical mountain pine beetle data used in this project. Thanks to Dr. Catherine Jarnevich and Dr. Jeff Morisette of the USGS Fort Collins Science Center for helpful advice on SAHM and climate models, and Dr. Paul Evangelista of NREL for his insights into MPB and forest ecology. I would also like to acknowledge all of the scientists who so graciously make their software and code publicly available and free to use; your contributions have made this research possible.

A number of colleagues and friends have been incredibly supportive and helpful throughout my studies. Tony Vorster has closely shared this journey with me and I thank him for all his support and a helpful review of this manuscript. Jared Stabach is a wizard of all things geospatial and has been very patient and generous in helping me with a variety of technical challenges. Thanks to all my other “science

buddies" in NREL and GDPE who have helped to make graduate school so rich and fulfilling and who have helped review my writing and troubleshoot my code. I have greatly enjoyed sharing this experience with all of you.

My family has been incredibly supportive throughout graduate school: thanks to Mom, Dad, and Daniel for the love and encouragement. To my love, Rachel Eisenstat: thank you for understanding the late nights and stress and for supporting me always, I hope it pays off for you! Lastly, I would like to extend my sincere gratitude to my dog, Bullwinkle, for the daily walks, laughs, and smiles.

PREFACE

Beginning in the late 1990s, a historical mountain pine beetle (MPB, *Dendroctonus ponderosae* Hopkins) outbreak in western North America resulted in widespread tree mortality as high as 90% in some lodgepole pine (*Pinus contorta*) stands. The Rocky Mountain region--Colorado, Utah, Wyoming, Montana, and Idaho--has previously experienced large MPB outbreaks, but the recent epidemic is unique in its scope and intensity. To better understand how environmental conditions in the region have changed over the past 50 years, and how these changes have contributed to the severity of the recent outbreak, I conducted a spatio-temporal analysis of historical and current MPB outbreaks in the Rocky Mountain region to identify the potential distribution of the beetle based on areas of environmental suitability. Additionally, I estimated future potential distributions in 2050 under two different climate change scenarios. Using MPB occurrence data from historical (1960-1980) and current (1997-2010) MPB infestations, and climate data from historical, current, and future time periods, I modeled the changes in both geographic and climatic space with three commonly used species distribution models (SDMs; MaxEnt, boosted regression trees, and generalized linear models) and a principal components analysis (PCA). Because predictions by SDMs are infrequently tested across time periods, I also evaluated the temporal transferability of the three models. The following four primary questions guided the research:

1. How has the potential distribution of MPB shifted under changing climatic conditions, and how is this reflected in the species' climatic niche; i.e., has it expanded, contracted, or shifted elevationally or latitudinally?
2. What were the primary drivers of MPB outbreaks from 1960-1980 and 1997-2010 in the Rocky Mountain region, and how have they changed?
3. How will the utilized climatic space of the beetle be expected to shift under projected future climatic conditions and how might this modify the distribution of the species?
4. Which SDM is most appropriate for predicting suitable habitat under future climate conditions?

Based on past work, I hypothesized that the climatic niche expanded between historical and current outbreaks, and that the resulting distribution expanded upwards in elevation across the study area. I anticipated that the current outbreak was driven by increased summer warming and lengthening of the growing season as opposed to the reduction of extreme cold temperatures. Lastly, I expected that the projected future conditions would restrict the potential distribution.

This thesis is the culmination of my graduate research and was written as a manuscript for submission to a peer-reviewed journal. Chapter one contains the manuscript in its entirety: an introduction and literature review, the methods used in the analyses, modeling and statistical results, a discussion of my findings, and conclusions. Chapter two discusses the key findings of my research and reflections on my research experience. The Appendices detail the data included in my database, the initial variables considered in the analysis, supplementary model results/maps, and different code (scripts) used in my analysis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

USING SPATIOTEMPORAL CORRELATIVE NICHE MODELS FOR EVALUATING THE EFFECTS OF CLIMATE CHANGE ON MOUNTAIN PINE BEETLE

Introduction

Global surface temperatures have warmed over the last three decades, with each successive decade warmer than the preceding decade (IPCC 2014). Substantial warming is projected in the U.S. Rocky Mountain region (Colorado, Utah, Wyoming, Montana, and Idaho) by mid-century with temperatures exceeding the global mean, and particularly pronounced at higher elevations (Bentz et al. 2010, Lukas and Gordon 2015). Combined with a shift in the timing and frequency of precipitation events, the Rocky Mountain region is forecast to grow hotter and more susceptible to drought in the coming decades (Seager et al. 2007, Lukas and Gordon 2015). These climatic changes portend significant ecological changes including species range shifts and an increase in landscape-shaping disturbances such as outbreaks of the mountain pine beetle (MPB, *Dendroctonus ponderosae* Hopkins), one of the principal drivers of landscape-level change in western North America (Dale et al. 2001, Parmesan 2006, Lenoir et al. 2008, Negrón and Fettig 2014).

The recent MPB epidemic is a historically large outbreak that has impacted over 6.5 million hectares of forest in the western U.S. (Bentz et al. 2010, USFS 2011). The MPB is a major disturbance agent that causes widespread tree mortality and substantially alters the structure, composition, and function of North American coniferous forests (Logan and Powell 2001, Carroll et al. 2006, Raffa et al. 2008). Given the severity of the recent beetle eruption, there has been a considerable focus on the ecology and long-term ramifications of the infestation on North American forests (Bentz et al. 2010, Negrón and Fettig 2014). The Rocky Mountain region has previously experienced large MPB outbreaks (Assal et al. 2014), but fire suppression, reduced habitat heterogeneity, and the climatic release of previously unsuitable habitats have driven an outbreak unique in its scope and intensity (Taylor and Carroll 2003, Carroll et al. 2006, Raffa et al. 2008).

The MPB prefers large-diameter trees and will infest any native pine in its range, though the two most important host species are lodgepole pine (*Pinus contorta*) and ponderosa pine (*P. ponderosa*) found in montane forests throughout the study area (Amman 1978, Logan and Powell 2001). The recent outbreak, which initiated in the mid-1990s, has also expanded into high-elevation subalpine forests (3,000 – 3,500 m) that were previously deemed too climatically harsh for eruptive MPB outbreaks (Logan and Powell 2001, Carroll et al. 2006). Potential hosts in subalpine forests include five-needle pines such as whitebark pine (*P. albicaulis*), limber pine (*P. flexilis*), and both Rocky Mountain and Intermountain bristlecone pines (*P. aristata* and *P. longaeva*, respectively) (Logan and Powell 2001). During an outbreak, MPB will overwhelm its host via a pheromone-driven "mass attack" that results in the establishment of egg galleries in the phloem (Negrón and Fettig 2014). The tree is killed by a blue-stain fungus introduced by the beetle that penetrates the xylem and blocks water transport from the soil to the canopy; however, phloem-feeding by adults and larvae also girdles the tree and contributes to its demise (Fairweather et al. 2013, Hubbard et al. 2013).

Outbreaks of MPB are greatly influenced by climate, which primarily influences MPB in three ways: through adaptive seasonality, cold-induced mortality, and drought stress on host trees (Creeden et al. 2014). Many stages of the beetle's life cycle are thermally regulated and MPB displays adaptive seasonality when the emergence of adults from host trees is synchronized to the climatic conditions at the appropriate time of year so that MPB experiences a one-year life cycle (Amman 1978, Safranyik 1978, Safranyik and Carroll 2006, Hicke et al. 2006, Sambaraju et al. 2012). Adaptive seasonality is conducive for large outbreaks, while maladaptive seasonality (two- or three-year life cycles) can restrict outbreak potential (Creeden et al. 2014). In contrast to adaptive seasonality, which facilitates large outbreaks, extreme cold temperatures may restrict the population success of MPB by reducing over-winter survival and causing widespread beetle mortality (Safranyik 1978, Campbell 2007, Sambaraju et al. 2012). Cold-induced mortality of over-wintering larvae is an important factor in MPB population dynamics, but MPB cold tolerance varies geographically and among seasons (Régnière and Bentz 2007). Drought indirectly drives outbreaks by restricting the host tree's ability to defend itself against beetle attacks and increases the probability of eruptive outbreaks (Safranyik 1978, Creeden et al. 2014). Drought is as an important component of beetle outbreaks, though many past studies have emphasized warming temperatures as the primary climatic driver behind the recent epidemic (Logan and Powell 2001, Hicke et al. 2006, Bentz et al. 2010, Jewett et al. 2011).

The relationship of climate to MPB has been modeled using a variety of statistical approaches, both mechanistic and correlative. Mechanistic, or process-based, models have been used to incorporate explicit relationships between climate and MPB performance to predict adaptive seasonality (Logan and Powell 2001, Hicke et al. 2006), cold-induced mortality (Régnière and Bentz 2007), and climatic suitability (Safranyik et al. 1975, Carroll et al. 2006, Bentz et al. 2010, Safranyik et al. 2010). Correlative models, which statistically correlate MPB outbreaks and climate, have been used to determine the climatic associations of MPB outbreaks and better understand the climatic conditions that support eruptive beetle outbreaks (Aukema et al. 2008, Evangelista et al. 2011, Jewett et al. 2011, Sambaraju et al. 2012, Creeden et al. 2014). Many of these models have been applied to future climate change scenarios to predict the climatic suitability for MPB outbreaks in a warming environment (Carroll et al. 2006, Hicke et al. 2006, Bentz et al. 2010, Safranyik et al. 2010, Evangelista et al. 2011, Sambaraju et al. 2012). While many of these studies evaluate MPB behavior through time, none explicitly compare the climatic drivers of historical outbreaks with the drivers of the recent epidemic in the Rocky Mountain region.

I utilized three correlative niche models to spatiotemporally evaluate the climatic correlates of MPB outbreaks since 1960. Correlative niche models—also known as bioclimatic envelopes, species distribution models, or ecological niche models—are probabilistic models that statistically correlate species' occurrences to its present environment and are often used to estimate a species' distribution and predict changes to the distribution under changing climatic conditions (Guisan and Zimmermann 2000). The technical foundations and relative performance of niche models have been widely reviewed (Guisan and Zimmermann 2000, Guisan and Thuiller 2005, Elith et al. 2006, Elith and Leathwick 2009), and these models have been implemented to explore the potential impacts of climate change on a variety of species (Thuiller et al. 2008, Monahan et al. 2013, Khanum et al. 2013, Anderson 2013).

I investigated how the climatic niche, potential distribution, and climatic drivers of MPB have changed across three time periods: 1960-1980 (historical), 1997-2010 (current), and 2040-2069 (future). I also refer to the potential distribution as climatically suitable habitat, or the abiotic conditions (topographic and climatic) that could support an MPB outbreak. Additionally, I tested niche models' transferability through time, or how well models project into different time periods with conditions not currently found in the study area. Four primary questions guided the research: (1) How has the potential distribution of MPB shifted under changing climatic conditions between historical and current outbreaks and how will this be expected to change under future climate change scenarios? (2) What were the primary climatic drivers of the historical and current outbreaks and how do they differ? (3) How will the utilized climatic space of the beetle be expected to shift under projected future climatic conditions and how might this modify the distribution of the species? and (4) Which correlative niche model is most appropriate for predicting suitable habitat under future climate conditions (i.e. temporal transferability)?

Data and Methods

The study was conducted in five U.S. states (Colorado, Idaho, Montana, Utah, and Wyoming) which have experienced, and continue to experience, extensive outbreaks (Figure 1, Appendix 1). To evaluate changes to the potential distribution and climatic drivers of MPB outbreaks, I used past and current U.S. Forest Service (USFS) aerial detection survey (ADS) data and a spatiotemporal modeling scheme that covered three time periods: 1960-1980 (historical), 1997-2010 (current), and 2040-2069 (future) (Figure 2). Additionally, I used a Principal Components Analysis (PCA) to show changes to the occupied climatic niche between historical and current outbreaks. Model transferability was assessed by training each niche model on historical data and projecting into current climate conditions, using current occurrence data as the test, or evaluation, data set (Figure 2).

Occurrence Data

The species occurrence data used in the analysis were generated from USFS ADS polygons that delineate the annual extents of MPB infestation and other forest disturbance across the five-state Rocky Mountain region (McConnell et al. 2000). Survey data were collected for “historical” (1960-1980) and “current” (1997-2010) time periods. All data were re-projected into the North American Datum 1983 (NAD83) Albers Equal Area projection to reduce latitudinal background selection of pseudo-absence (background) points in the niche models (Brown 2014). All MPB polygons from each study period were dissolved into a single layer and a sample of 5,000 stratified random points (where strata are polygons) was generated from within this layer using the *Geospatial Modelling Environment* software (Beyer 2012). This sample of occurrence localities was spatially filtered with the *SDMToolbox* so that no occurrence localities were within 10 km of another occurrence (Brown 2014). Spatial filtering can reduce model overfitting and spatial autocorrelation, and ensures independence of the test and training data when using a cross-validation evaluation technique (Veloz 2009, de Oliveira et al. 2014, Boria et al. 2014, Radosavljevic and Anderson 2014). I filtered at 10 km because MPB generally occurs in mountainous terrain with high spatial heterogeneity, similar to previous modeling studies that used the 10-km filter in mountainous regions (Pearson et al. 2007, Anderson and Raza 2010, Boria et al. 2014). Spatial filtering reduced the historical dataset from 5,000 original points to 882 points. The current data exhibited a wider geographic spread, and therefore contained more points, so the current data were also reduced to 882 points via random point selection to maintain consistency across the time periods.

Historical data were acquired from individual USFS Regional Offices. The historical data were originally collected on marked topographic quadrangles and georeferenced and digitized in geographic information systems (ArcGIS, ESRI v10.2). Because there are historical surveys that remain un-digitized, and therefore unavailable for use in this study, it should be noted that the historical dataset is partially incomplete and may not reflect the full range of MPB presence in the years 1960-1980. Current data were downloaded from the Insect and Disease Detection Survey (IDS) Data Explorer (USFS 2014).

Climate data

Climate data were acquired from ClimateWNA (version 5.10) at 1-km grid cell resolution (Wang et al. 2012). Historical climate data were selected for the 30-year normal period spanning 1951-1980, and current climate data span 1981-2010. The 30-year normal for future climate projections covers the years 2040–69, which I refer to as “2050”. I used the global mean of 15 GCMs for two representative concentration pathways (RCP), RCP 4.5 and RCP 8.5, selected from phase five of the Coupled Model Intercomparison Project (CMIP5) multi-model data set that corresponds with the Fifth Assessment Report from the Intergovernmental Panel on Climate Change (IPCC) (Moss et al. 2010, Taylor et al. 2012, Wang et al. 2012). RCP 4.5 is considered a medium stabilization scenario (~650 ppm CO_2 equivalent by 2100) that encompasses the vast majority of the scenarios assessed in the Fourth Assessment Report (van Vuuren et al. 2011). RCP 8.5 is considered a very high emissions scenario (~1370 ppm CO_2 equivalent by 2100) that assumes no current or future climate policy (van Vuuren et al. 2011).

Forty-five initial variables were chosen from the full ClimateWNA dataset (Appendix 2, Table 5) based on the known climatic and environmental influences on MPB biology and ecology. These variables were tested for correlation based on the Pearson, Spearman, and Kendall coefficients and highly correlated variables ($|r| \geq 0.7$) were filtered using expert knowledge of MPB ecology and were chosen to represent seasonal climatic influences on MPB. However, four pairs of highly correlated variables were retained in the final analysis to examine seasonal influences on the beetle. The final predictors included 14 climatic and topographic variables (Table 1).

Spatiotemporal Modeling

I used three correlative niche models (Appendix 3) and a PCA to evaluate historical and current MPB outbreaks. The niche models were used to estimate the potential distribution of MPB and the PCA was used to evaluate the potential climatic niche shift in multidimensional space. Three distinct models were run for this analysis: maximum entropy (MaxEnt) (Phillips et al. 2006), boosted regression trees (BRT)

(Elith et al. 2008), and generalized linear models (GLM) (McCullagh and Nelder 1989, Austin 2002); see details on model parameterization in Appendix 3. These three models have consistently demonstrated high performance across species functional groups and compare favorably to other correlative models (Elith et al. 2006, Guisan et al. 2007, Austin 2007, Stohlgren et al. 2010). All models were trained using the same 14 variables across all time periods (Table 1). Each model was tested internally using a 10-fold cross-validation (Fielding and Bell 1997). The potential geographic overlap between models was calculated with Schoener's *D* statistic via the 'ENMeval' package in R v.3.1.2 (Warren et al. 2008, Muscarella et al. 2014, R Core Team 2014). All final maps were clipped to the combined forest classifications from the National Land Cover Dataset (NLCD) for 2001, 2006, and 2011 (Homer et al. 2007, Fry et al. 2011, Jin et al. 2013). Forested areas include the sum of forest land cover classification codes 41 (deciduous forest), 42 (evergreen forest) and 43 (mixed forest).

MaxEnt is a general-purpose machine learning method that was run in the stand-alone software package (Phillips et al. 2006). A number of recent studies have underscored the importance of carefully calibrating the MaxEnt model (Merow et al. 2013, Shcheglovitova and Anderson 2013, Radosavljevic and Anderson 2014). To parameterize MaxEnt models for MPB I experimentally tuned the parameters using the "ENMeval" package in R v.3.1.2 (Appendix 3, Figure 7, 8; Muscarella et al. 2014, R Core Team 2014). I ran all MaxEnt models (historical, current, and projected) using the "all features" setting, a regularization multiplier of 3.0, and 20,000 background samples. Based on the ENMeval metrics, these settings produced the best performing models with biologically reasonable response curves (Appendix 4, 5).

Boosted regression trees are an ensemble method for fitting statistical models that use regression trees and boosting to combine many simple models and improve performance (De'ath 2007, Elith et al. 2008). Boosted regression trees tend to overfit models, so regularization methods are used to constrain the fitting procedure by optimizing three parameters: number of trees, learning rate, and tree complexity (Elith et al. 2008). The BRT models were fitted using the Software for Assisted Habitat Modeling (SAHM) and I experimentally parameterized the learning rate and tree complexity to derive models with at least 1,000 trees and biologically sensible response curves (Morisette et al. 2013). The best settings that resulted in at least 1,000 trees for the historical model had a learning rate of 0.005 and tree complexity of 5; the current model was parameterized at 0.005 and 3.

Generalized linear models are a regression approach that fits parametric terms using some combination of linear, quadratic, and/or cubic terms (Elith et al. 2006). Within SAHM I fit the GLM to a binomial distribution with a logit link function, and the SAHM algorithm selected the optimal model based on a bidirectional stepwise procedure to select covariates based on Akaike's Information Criterion (AIC; Morisette et al. 2013).

I evaluated elevational range shifts, range expansion, and range contraction to assess geographic trends across time periods (Figure 2). In addition to calculating these values for each individual model, I created an ensemble prediction for each time period to assess the average progression through time. Ensemble models are a solution to inter-model variation and capture the areas of agreement across models (Araújo and New 2007). To create the ensemble, binary suitability maps were produced using a fixed 95% sensitivity threshold; that is, the threshold was the lowest predicted probability that encompassed 95% of the occurrence localities (Peterson et al. 2011). The binary maps for each model were combined so that the resulting ensemble map contained only pixels that were deemed environmentally suitable by all three models (Stohlgren et al. 2010).

The correlative niche models encompass two strategies for modeling presence-only data. MaxEnt draws pseudo-absences from a random sample of background pixels to account for the presence-only structure of the occurrence data, whereas the BRT and GLM models are derived from regression techniques generally associated with presence-absence data (Phillips et al. 2006). Because absence data were not available for the historical period, I used background data as pseudo-absences for the BRT and GLM models (Phillips et al. 2009). Background samples were constrained by a kernel density estimator

(KDE) to account for potential sampling bias that may exist when aerial surveys are primarily flown over federal lands (Kumar et al. 2014a, b). The KDE restricted background sampling to general “use areas” for MPB so that all background samples were drawn from environmental conditions the species is more likely to reach (Merow et al. 2013). The constrained background sampling corrects for sampling bias in the species occurrences by applying the same bias to the background points, thereby cancelling out the bias in the modeling process (Phillips et al. 2009). The KDE was generated in SAHM using a 95% isopleth on MPB occurrence data; i.e., the resulting mask represented the smallest area providing a 95% probability of finding MPB (Fieberg 2007, Morisette et al. 2013). I created separate surfaces for both historical and current occurrence data that were used to restrict the background “absences” in BRT and GLM; these surfaces were used as a bias file in MaxEnt.

Models were evaluated using a threshold independent metric, AUC, and a threshold dependent metric, sensitivity (the true positive rate). The AUC metric is a commonly used statistic that represents an overall measure of a model’s predictive accuracy and summarizes the model’s ability to distinguish between a species’ presence and absence (Peterson et al. 2011). Though AUC can be a misleading measure of model performance, it was useful for this study because all models were trained on the same geographic extent and background samples were extracted from the general use area defined by the KDE (Lobo et al. 2008). I also evaluated sensitivity to assess model performance through time when climatic conditions may differ. Sensitivity is the rate of known presences correctly predicted by the model prediction (1 – the omission error rate) and represents the absence of omission error (Peterson et al. 2011).

To assess the temporal transferability of the various modeling techniques, I trained each model on the historical data (historical occurrences and climate data) and projected them onto current climate conditions (Figure 2). To assess the quality of the predictions of the forecast model, I tested the predictions—trained with historical occurrence data—against current occurrence localities and generated AUC and sensitivity statistics for each model. Sensitivity was calculated by thresholding the projection at the same 95% sensitivity threshold of the historical model; for example, if the 95% threshold for the historical model was 0.26 then this value was used as the threshold for the projected model as well. I calculated AUC values for the projected models using the ROC/AUC calculator (Schroeder 2006), and created multivariate environmental similarity surface (MESS) maps to quantify the extent of extrapolation in model projections (Elith et al. 2010). The MESS maps were generated within SAHM (Morisette et al. 2013). All current models were then projected to two climate scenarios for 2050, RCP 4.5 and RCP 8.5. The models were trained using the current occurrence localities and climate data, and forecast climate conditions were substituted to provide a projection of future climatic suitability.

Lastly, I used a PCA model in R v.3.1.2, adapted from Broennimann et al. (2012), to assess potential shifts of the climatic niche in multivariate environmental space (R Core Team 2014). I ran three separate PCA analyses with all 14 environmental variables to contrast the fundamental niche shift of MPB across time periods: historical to current, current to RCP 4.5, and current to RCP 8.5. To prepare the data for the PCA, 20,000 random background points were selected from across the study extent and variable values were extracted at each point. Additionally, data were extracted at each of the 882 occurrence localities for each time period. Contrasting principal components were overlaid to determine the extent of MPB in ordinal space and to assess the niche overlap between time periods (West et al. 2015). Additionally, I calculated niche overlap in climatic space using Schoener's *D* metric, which varies from 0 (no overlap) to 1 (complete overlap) (Warren et al. 2008, Broennimann et al. 2012).

Results

Historical models

MaxEnt and BRT were the top performing historical models with test AUC values of 0.85; GLM slightly underperformed these two models with a test AUC of 0.81 (Table 2). All models displayed a good fit, meaning they captured a large fraction of the total variability in the data, with minimal difference between training and test AUC values for the MaxEnt and GLM models. A pairwise comparison of niche overlap predicted by the models revealed strong agreement among the models. MaxEnt and BRT shared 85% overlap (as calculated by Schoener's *D* statistic), MaxEnt and GLM had 80% overlap, and GLM and BRT also had 80% overlap. The historical ensemble model predicted 249,002 km^2 of climatically suitable habitat for MPB from 1960-1980 (Table 3).

The top three predictors for the MaxEnt model were (in order of importance) summer precipitation, summer degree-days above 18°C, and precipitation as snow between August and the following July (Table 2). The top predictors for BRT were summer degree-days above 18°C, precipitation as snow, and the climatic moisture deficit. The top predictors for GLM were summer degree-days above 18°C, winter degree-days below 0°C, and elevation. The only top predictor common to all three models was summer degree-days above 18°C.

Current models

Similar to the historical models, MaxEnt and BRT were the top performing current models with a test AUC value of 0.82; the current GLM model had an AUC of 0.80 (Table 2). Again, all models produced strong predictions, though with a slight decrease in performance across the board, and all models also showed good fit with low ΔAUC values. A comparison of the niche overlap between current model predictions again showed high agreement among the models. MaxEnt and BRT shared 86% overlap, MaxEnt and GLM had 82% overlap, and GLM and BRT also had 84% overlap.

The top predictors for the current outbreak showed more consistency among the models than for the historical models. Summer degree-days above 18°C and the climatic moisture deficit were top predictors in all three models. The other top predictors included summer precipitation (MaxEnt), slope (BRT), and the date of the end of the frost-free period (GLM).

All models estimated a substantial range expansion for the pine beetle between the historical and the current time periods (Table 3). Among individual models, the BRT predicted the highest net expansion with the current model estimating 38,444 km^2 of additional climatically suitable habitat (Table 3). MaxEnt showed a net expansion of 35,577 km^2 in the current time period and GLM estimated an expansion of 22,929 km^2. The current ensemble estimates that forests in the Rocky Mountain region contain 295,207 km^2 of climatically suitable habitat for MPB from 1997-2010, an estimated net expansion from the historical ensemble of 46,205 km^2 (Table 3).

In addition to an overall range expansion, the model results suggest that this range expansion correlates with an upward shift in elevation (Figure 3, Table 3). All of the individual models show a statistically significant upward shift in the mean elevation across the potential distribution. Again, the ensemble models demonstrate the greatest change with an elevation shift of +115 meters ($p < 0.0001$). Of the individual models, the BRT showed the greatest shift at +99 m ($p < 0.0001$), which corresponds with the greatest net expansion across the range. The MaxEnt models exhibited a +79 m ($p < 0.0001$) shift and GLM showed an increase of +22 m ($p < 0.0001$) between the two time periods.

Future projections

Under the RCP 4.5 scenario, all models predicted a net contraction of climatically suitable habitat for MPB (Figure 4, Table 3). The GLM model predicted the greatest contraction with a decrease of 81,553 km^2. MaxEnt predicted a contraction of 43,082 km^2 and BRT estimated a 32,335 km^2 decrease in climatically suitable habitat. The ensemble map of the three RCP 4.5 forecasts estimated 228,111 km^2 of climatically suitable habitat in 2050, a contraction of 67,096 km^2 from the current estimation and less than historical ensemble as well. The projections under RCP 4.5 showed greater disagreement than the historical and current models. The MaxEnt and BRT models showed a predicted niche overlap of 78%; MaxEnt and GLM overlap by 87%; and the BRT and GLM models overlap by 75%.

The second forecast projected the models onto data from the RCP 8.5 scenario. The patterns of contraction seen in the RCP 4.5 projections held true for RCP 8.5 forecasts as well: GLM predicted the greatest contraction of 116,990 km^2, MaxEnt predicted the next largest contraction at 67,404 km^2, and

BRT estimated a reduction of 45,486 km^2 of suitable area. The ensemble map of the RCP 8.5 forecasts estimates 194,420 km^2 of climatically suitable habitat for 2050, a contraction of 100,787 km^2 from the current estimation and a smaller range than under historical conditions. The RCP 8.5 projections had the least agreement of all the temporal segments. MaxEnt and BRT models showed a predicted niche overlap of 74%; MaxEnt and GLM overlap by 86%; and the BRT and GLM models overlap by 70%.

Model Transferability

All three historical models demonstrated good fit and high AUC values when projected into the current climate conditions; based on the test AUC values, each model performed better than the current models trained on the current occurrences (Table 4). Based on AUC, the MaxEnt and BRT models were the top performing models with AUC of 0.87, and GLM had an AUC of 0.86. MaxEnt had a slightly higher sensitivity of 82% compared to BRT (81%), but both were lower than GLM, which correctly predicted 90% of the current occurrences. Overall, all three models provided reasonable predictions across time periods.

While evaluating model transferability, I used MESS maps to track the extent of extrapolation in model projections (Figure 5, Elith et al. 2010). The MESS maps show minimal extrapolation in all projections, from historical to current and current to 2050 (RCP 4.5 and RCP 8.5). Areas of high extrapolation were generally outside the estimated climatic niche of MPB, in alpine environments or the southern reaches of the study extent dominated by non-forested grassland, shrubland, and desert.

Potential Niche Shift

I evaluated shifts in the climatic niche space utilized by the mountain pine beetle across time periods using the PCA. The PCA shows the relative niche occupancy along each axis of the PCA. The first PCA compared the historical climatic niche to the current niche and was run with all 14 climatic and topographic variables and showed a significant niche between outbreaks (Figure 6). Three variables displayed similar loadings for the first principal component (PC1): winter degree-days below 0°C, beginning of the frost free period, and spring degree-days below 0°C. The highest loading of the second component (PC2) was autumn precipitation. The niche overlap (Schoener's *D*) between the two time periods was 0.30. This suggests that only 30% of the ordinal historical niche was utilized by the species during the current outbreak. This shift shows that the historical and current climatic niche were not significantly similar (p = 0.207).

I ran the PCA comparing the current climatic niche with the potential niche under two future climate change scenarios, RCP 4.5 and RCP 8.5 under the assumption that the current occurrence localities would remain suitable habitat under future conditions (Figure 6). The results were similar under both future scenarios. The top loadings of the PC1 were winter degree-days below 0° C, beginning of the frost free period, and spring degree-days below 0° C. The PC2 was loaded primarily by autumn and spring precipitation. The niche shift was slightly more pronounced under RCP 8.5, which shared 54% of the ordinal climate space with the current niche. The overlap between the current niche and RCP 4.5 was 61%.

Discussion

The models used in this study represent approximations of climatic suitability for MPB outbreaks and shifts in suitable area is estimated based on the correlative relationships between the predictors and the occurrence localities. The model results should be treated as distributional hypotheses that are limited to the predictors, extent of the study region, and location of MPB occurrences (Lobo et al. 2008). My results imply that climatic changes in the latter half of the 20th century significantly increased the amount of climatically suitable habitat for MPB in the U.S. Rocky Mountain region and that the recent MPB outbreak displayed a different climatic signature than historical outbreaks. The expansion of climatically suitable habitat reflects an upward elevational shift into previously unsuitable habitats and a change in MPB's climatic niche. Yet, despite the recent expansion of suitable habitat for MPB, future projections suggest that climate warming will reduce the amount of climatically suitable areas by mid-century.

Climatic drivers of mountain pine beetle outbreak and range expansion

My results revealed both direct and indirect climatic drivers of MPB outbreaks. The primary climatic drivers for both the historical and current outbreaks were summer heat accumulation and drought (Table 2), which align with past findings on the climatic influence on MPB outbreaks (Bentz et al. 2010, Evangelista et al. 2011, Chapman et al. 2012, Creeden et al. 2014). However, my model results showed different climatic signatures between historical outbreaks and the recent epidemic. All three correlative niche models agree that the climatic moisture deficit was the most important predictor variable for the current outbreak which suggests that drought has played a larger role in the current outbreak than in historical outbreaks. An increase in summer heat, particularly at higher elevations, has certainly contributed to the intensity of the recent outbreak by facilitating adaptive seasonality and reducing the risk of over-winter mortality, but long-term drought that has contributed to increased tree mortality in the region has also made host trees far more susceptible to eruptive MPB outbreaks (Hicke et al. 2006, van Mantgem et al. 2009). My results indicate that drier summers with reduced moisture availability and an earlier spring that diminishes snowpack are critical elements of shifting MPB distributions that have intensified over the past 50-60 years in the region (Westerling et al. 2006, Bentz et al. 2010).

These climatic drivers resulted in a substantial expansion of the climatically suitable habitat of MPB between 1960 and 2010. Though there was some variability among the models, all three models, as well as the ensemble model, showed a net expansion of suitable habitat during the current outbreak (Table 3). As conditions grew warmer over the past 50 years, MPB expanded into previously unsuitable high-elevation forests (Carroll et al. 2006), which is reflected in the primary habitat gains along the range margins and an increase in the average elevational range of the species (Table 3). A considerable portion of this expansion occurred in northwest Wyoming in the Greater Yellowstone Ecosystem. This ecosystem has recently experienced high rates of whitebark pine mortality driven by warmer, drier conditions (Jewett et al. 2011). These climatic conditions correspond with the primary climate variables that drove the expansion of suitable environment for MPB throughout the region, which supports past arguments that climate change is largely responsible for the expansion of the beetle into this previously unsuitable domain (Logan and Powell 2001, Carroll et al. 2006). My estimates of the current expanse of suitable environment are similar to that of Evangelista et al. (2011); however, by using climate data through 2010 I was able to capture suitable habitat in northwest Wyoming that was not predicted by their models using climate data through 2000.

The transition of MPB into high-elevation forests is also shown in the utilized climatic niche. Three predictors contained a majority of the variability in the first principal component: the beginning of the frost-free period (bFFP) and degree-days below 0°C in both the spring and winter; the shift in the second principal component was driven by increases in precipitation in the spring and autumn. The climatic niche of the current outbreak shifted positively along both axes of the PCA, which indicates higher correlation with the principal loadings of the axes. Higher elevations would be expected to have a later last frost, more cold days in the winter and spring, and more precipitation in the spring and fall, and the positive correlations of these variables with MPB occupancy in the current outbreak suggest that the current outbreak occupied suitable habitats at higher elevations than in the historical outbreak.

With regards to future predictions of climatic suitability, my models projected a net contraction under both future scenarios, RCP 4.5 and RCP 8.5. The net contraction was more pronounced under RCP 8.5, the high emissions scenario, but both projections were indicative of a decrease in climatically suitable habitat for MPB. There are a number of possible explanations for this trend, though none were tested explicitly in the modeling. The life cycle of MPB is under direct temperature control and population success is closely tied to phenology; adult beetles must emerge late enough in the summer to avoid lethal freezing, but not so late as to reduce ovipositional potential through fall/winter cooling (Logan and Bentz 1999). Projected decreases in suitable habitat are likely related to a reduction in areas of adaptive seasonality; i.e., conditions that promote earlier emergence may expose overwintering larvae to cold temperatures before they are sufficiently cold-hardened (Hicke et al. 2006). Further warming could also disrupt current suitable habitat by promoting maladaptive seasonality or disrupting the beetle’s physiology (e.g. flight), which could reduce the effectiveness of the species' "mass attack" strategy and other key life stages (McCambridge 1971, Safranyik 1978, Logan and Bentz 1999). Though climate change is expected to intensify all aspects of insect outbreaks, warming at lower elevations and latitudes could result in the reduction of suitable environments for MPB, which is shown in the model predictions (Figure 4, Logan et al. 2003). There is less confidence in forecasts of precipitation in climate models, so anticipating the effects of drought on climatically suitable habitat in the future may be more difficult than linking potential changes to warming temperatures.

The PCA revealed a potential niche shift under both RCP 4.5 and RCP 8.5 (Figure 6). Without known future occurrences I was only able to estimate the background environment and extract forecast conditions at the current outbreak localities, so this approach assumes that the current suitable habitat will also be suitable—biologically and climatically—in the future. Moreover, this approach does not take into account any future expansion, biotic interactions, and currently unaccounted localities. The two principal components were loaded similarly to the historical/current PCA; the first component reflected the beginning of the frost-free period (bFFP) and degree-days below 0°C in both the spring and winter and the second principal component was loaded by precipitation in the spring and autumn. The future niche space was similar under both climate scenarios, but expansion of the niche under RCP 8.5 was slightly more pronounced than RCP 4.5, which would be expected because it is a more severe forecast. Overall, the future climatic space shifted negatively along the x-axis and positively along the y-axis, which suggests a reduction in degree-days below 0°C in the winter and a warmer, earlier spring. The shift along the y-axis indicates an increase in precipitation in the spring and fall, though this was fairly minimal compared to the horizontal shift. The PCA suggests that currently occupied habitats will continue to grow warmer and that the high-elevation habitats will become more conducive to beetle outbreaks.

Spatiotemporal model transferability

Predicting a species' response to climate change assumes that models are transferable through time and that models adequately extrapolate to novel conditions, or conditions not currently found in the study area. Predicting species' responses to novel conditions often involves extrapolation beyond the range of the data used to train the model, which can be more complicated than interpolative forecasting because temporal or spatially independent data is often unavailable to test model predictions (Williams et al. 2007). This transferability (also called "generality") refers to a model's ability to make useful predictions in a different context from which it was trained, and models with better transferability would be expected to make more useful predictions (Dobrowski et al. 2011). In general, broadly applicable models provide more useful predictions than those that only accurately predict occurrence based on a narrow set of conditions (Wenger and Olden 2012).

Multiple studies have addressed the issue of temporal transferability for a range of models (Araújo et al. 2005, Pearman et al. 2008, Kharouba et al. 2009, Dobrowski et al. 2011, Heikkinen et al. 2012), but given the relative lack of temporally independent data sets such investigations are still fairly uncommon (Araújo et al. 2005). Because a species' observed distribution alone cannot provide information on how a species may respond to novel conditions, assessments of temporal transferability are important for determining the usefulness of predicted responses to climate change (Fitzpatrick and Hargrove 2009). In my study, all three model projections provided reasonably good predictions (test AUC values > 0.85) when projected through time and there was little difference in model performance (Table 4). Given past research on transferability, the relative similarity between model projections was expected; in general, the functional traits of species influence transferability more than differences in the modeling algorithms (Kharouba et al. 2009, Dobrowski et al. 2011, Heikkinen et al. 2012). The results from this study may be useful for predicting the climate change responses of other native bark beetles (Coleoptera: Curculionidae, Scolytinae) such as the spruce beetle (*D. rufipennis* Kirby) and western pine beetle (*D. brevicomis*).

The choice of modeling algorithm for forecasting will largely be determined by the goals of the project, but my analysis suggests that a simpler model, such as the GLM, may be more appropriate for future predictions that seek to limit omission error. GLM performed 8-9% better than the BRT and MaxEnt models but did not adequately discriminate between unsuitable high-elevation environments and the mid-elevation environments that are the primary habitats of the beetle. Because of this generality the GLM predicted a much narrower elevational shift and less expansion of suitable habitat between the time periods; however, the generalized prediction yielded more accurate predictions of current outbreaks. The GLM had the lowest omission error, which is especially important in analyses of relocation, translocation, or species reintroduction, as well in assessments of risk from invasive species or disease (Araújo and Peterson 2012).

The MESS maps reveal that, despite projection across temporal domains, extrapolation in the model projections was fairly limited (Figure 5). None of the forecasts from either the historical to current time period, or from the current to future scenarios, exhibit significant novelty in regards to the variables used in the models, and regions that did exhibit novel conditions are not generally susceptible to MPB outbreaks (non-forested, high-elevation alpine and southern shrub and desert ecosystems). There are a couple of possible explanations for this. First, the chosen time periods may not be separated by enough time to show significant climatic changes. Yet, the past three decades have shown unprecedented warming, a trend that is anticipated to continue over the next three decades (IPCC 2014). Both current and future climate data should reflect this warming, and novel conditions would be expected. Instead, it is more likely that the projections lacked novel conditions because the models were trained on data drawn from a heterogeneous landscape. Rocky Mountain landscapes are highly varied and have significant topographic relief throughout the region in addition to a large latitudinal gradient. As a result, though certain locations might see drastic climatic changes, the new conditions are likely found elsewhere in the study area and were used to train the model. For MPB, a generalist herbivore, estimates of climatically suitable areas in the future can be viewed with a higher degree of confidence than a specialist species with a more restrictive elevation or latitudinal range because model projections were not extrapolating to novel conditions.

Modeling future suitability requires a number of assumptions that may not be true under novel climatic and environmental conditions. For example, over the past 20-30 years the study region has undergone significant population growth in exurban areas that overlap with MPB habitat, and the current outbreak largely coincides with an increase in large forest fires (>400 ha) across the same habitats (Westerling et al. 2006, Maestas et al. 2011). These changes have introduced substantial environmental change to habitats that support MPB, and as a result, current MPB occurrences may not reflect a species at equilibrium with its environment, which is one of the key assumptions of correlative niche models (Wiens et al. 2009). This can be problematic when applying modeling algorithms to novel temporal domains under future climate scenarios, and future predictions should be interpreted cautiously (Araújo and Peterson 2012). Correlative niche models are also unable to account for evolutionary adaptations that may occur over time (Pearson and Dawson 2003). When projecting future responses of MPB to climate change, I can estimate future suitable habitat but cannot forecast the effects of warming on host trees or how the beetle may respond to other rapidly changing environmental conditions (Bentz et al. 2010). I have high confidence in the modeled response of the beetle to 20th-century warming because the predictions are rooted in actual occurrences, but future projections should be interpreted cautiously.

Furthermore, there is inherent uncertainty in the data used in this analysis. Although improvements in global positioning systems (GPS), GIS, and aerial detection techniques have reduced the uncertainty of recent outbreak polygons, rates of omission—when a category other than ‘no damage’ is found on the ground but no observation was recorded on the aerial survey map—for aerial detection survey data can be as high as 35% in lodgepole pine forests. The historical MPB dataset may have higher error rates resulting from georeferencing and digitizing old topographic quadrangles (Johnson and Ross 2008). For this analysis, I can reasonably expect that a 1-km pixel would encompass most of the uncertainty from the aerial survey; however, this geographic error may result in an MPB occurrence correlating with different conditions than the species experienced in the environment. There are also varying levels of uncertainty associated with the different climate data products used. The various interpolative and downscaling techniques used by ClimateWNA introduce uncertainty into the data, and future climate forecasts retain internal model variability (Beaumont et al. 2007, Wang et al. 2012). Through careful calibration and a deliberate consideration of this uncertainty I was able to reduce some of the uncertainty in my modeling, but model predictions—particularly forecasts into future domains—should be interpreted as estimates and geographic approximations, not certainties.

Conclusions

My research reveals a significant expansion of climatically suitable area for MPB over the past half-century in both geographic and climatic space; however, projected warming may reduce climatic suitability under future climate scenarios. Furthermore, my results suggest that the recent MPB epidemic showed a different climatic signature than historical outbreaks as drought drove model predictions more so than temperature increases. The shift of climatically suitable habitats into higher elevations is expected to continue in the future, and this shift threatens sensitive high-elevation ecosystems such as those dominated by whitebark pine, but may also reflect the de-stabilization of currently suitable habitats at lower elevations (Jewett et al. 2011). My results confirm that climate change has driven a range expansion of MPB and corroborates past research on the effects of climate on the spatial distribution of MPB outbreaks.

I have also demonstrated that three common correlative niche models provide fairly reliable estimates of species response to climate change. While studies utilizing correlative models should always be aware of the assumptions and limitations of the models, correlative niche models can be an effective and reliable tool in predicting change across temporal domains (Pearson and Dawson 2003, Araújo and Peterson 2012). Simpler algorithms, like the GLM, may provide more general predictions that project better across temporal domains and reduce omission error.

Acknowledgements

Funding and support for AMS was provided by the Natural Resource Ecology Laboratory and the Warner College of Natural Resources at Colorado State University and through the Neil B. Kindig Fellowship of the Colorado Mountain Club. SK was jointly funded by the U.S. Geological Survey and NASA. I would like to thank Tim Assal (U.S. Geological Survey, Graduate Degree Program in Ecology, CSU), Brian Howell and Justin Backsen (Forest Health Protection, USFS, Rocky Mountain Region), and Dick Halsey (Forest Health Protection, USFS, Boise Field Office) for their assistance in procuring historical MPB data. I thank Tony Vorster for a helpful review of the manuscript. Lastly, I would like to acknowledge all of the scientists who so graciously make their software and code publicly available and free to use; your contributions have made this research possible.

Table 1: Predictor variables used in the three niche models. For a detailed description of climate variables see Wang et al. (2012).

Variable	Description	Rationale
CMD	Hargreaves climatic moisture deficit (CMD). Sum of the monthly difference between reference atmospheric evaporative demand (E_{ref}) and precipitation. A higher CMD reflects a greater moisture deficit.	Drought affects the host tree's ability to defend itself against bark beetle attack (Safranyik 1978, Creeden et al. 2014). Below-average precipitation across the growing season correlate with increased MPB (Amman 1978, Carroll et al. 2006).
PAS	Precipitation as snow (PAS, mm) between August of previous year and July of current year	
PPT_sp	Spring precipitation between March - May	
PPT_sm	Summer precipitation between June - August	
PPT_at	Autumn precipitation between September - November	Reduction in autumn moisture immediately following an attack benefits larval over-winter survival (Amman 1978).
bFFP	Julian date on which the frost-free period (FFP) begins	Spring temperature affects larval development (Amman 1978, Aukema et al. 2008)
eFFP	Julian date on which the frost-free period ends	Early onset of frost period in the late summer and autumn may affect egg and larval development (Safranyik 1978)
Tmin_wt	Winter mean minimum temperature (℃)	Severe winter temperatures can reduce over-winter survival and cause widespread beetle mortality (Safranyik 1978, Campbell 2007, Sambaraju et al. 2012)
DD_0_wt	Winter degree-days below 0° C	
DD_0_sp	Spring degree-days below 0° C	Spring temperature affects larval development (Amman 1978, Aukema et al. 2008)
DD18_sm	Summer degree-days above 18° C	Summer heat accumulation affects many aspects of the MPB life cycle including emergence, flight, and egg hatch (Sambaraju et al. 2012)
elevation	Digital elevation model (DEM) at 1-km resolution	Topographic variables roughly define suitable topography for host species (Safranyik 1978, Sambaraju et al. 2012)
slope	Maximum change in elevation between each cell and its eight neighbors	
aspect	Downslope direction of a grid cell	

Table 2: Model summary and results. Training AUC calculated with an internal 10-fold cross-validation; Δ AUC is a measure of model overfit (over parameterization) and lower values indicate better fit.

Model Description			*Model Evaluation*		
Model	**Number variables***	**Top variables**	**Training AUC**	**Test AUC**	**Δ AUC**
MaxEnt – Historical	12	Summer precipitation, summer degree-days (18°C), precipitation as snow	0.86	0.85	0.01
BRT – Historical	10	Summer degree-days (18°C), precipitation as snow, climatic moisture deficit	0.88	0.85	0.03
GLM - Historical	6	Summer degree-days (18°C), winter degree-days below 0°C, elevation	0.81	0.81	0
MaxEnt - Current	12	Climatic moisture deficit, summer precipitation, summer degree-days (18°C)	0.82	0.82	0
BRT - Current	6	Climatic moisture deficit, summer degree-days (18°C), slope	0.84	0.82	0.02
GLM – Current	7	Climatic moisture deficit, summer degree-days (18°C), end of frost-free period	0.8	0.8	0

* All variables were included in the initial run, but were reduced through a jackknife test of variable importance. In MaxEnt, variables of low importance were manually removed. The BRT and GLM algorithms in SAHM automatically removed variables of low importance.

Table 3: The predicted area (km^2) of climatically suitable habitat for the mountain pine beetle across historical, current, and future time periods. Predicted changes in area for current and 2050 estimates show the calculated areas of suitable habitat and the extent of expansion and contraction from the preceding time period (i.e. range expansion for the current estimates reflects change compared to the historical predictions). The elevation shift reflects the mean elevation of climatically suitable habitats. All values reflect suitable habitat clipped to the NLCD forest layers.

Model	Total Area	Range expansion	Range contraction	Net	Elevation shift (m)
Historical					
MaxEnt - Historical	275,565	-	-	-	-
BRT - Historical	267,840	-	-	-	-
GLM - Historical	311,565	-	-	-	-
Ensemble - Historical	249,002	-	-	-	-
Current					
MaxEnt - Current	311,142	41,254	5,677	35,577	+ 79
BRT - Current	306,284	45,510	7,066	38,444	+ 99
GLM - Current	322,123	17,224	6,666	10,558	+ 22
Ensemble - Current	295,207	52,350	6,145	46,205	+ 115
Future - RCP 4.5					
MaxEnt – RCP 4.5	267,970	2,987	46,069	-43,082	+ 41
BRT - RCP 4.5	273,949	3,830	36,165	-32,335	+ 19
GLM - RCP 4.5	240,570	46	81,599	-81,553	+ 110
Ensemble - RCP 4.5	228,111	1,570	68,666	-67,096	+ 87
Future - RCP 8.5					
MaxEnt - RCP 8.5	243,738	1,440	68,844	-67,404	+ 74
BRT - RCP 8.5	260,798	2,381	47,867	-45,486	+ 24
GLM - RCP 8.5	205,133	0	116,990	-116,990	+ 171
Ensemble - RCP 8.5	194,420	731	101,518	-100,787	+ 139

Table 4: Evaluation of model transferability from historical to current climate conditions. Sensitivity is based on the 95% sensitivity threshold used for the historical model and applied to projections with the current climate data. The current occurrence localities were used as test data, temporally independent from the training data. Sensitivity is the number of correctly predicted current occurrences out of 882 occurrence localities.

Model	Test AUC	Sensitivity
MaxEnt – historical projected to current	0.87	82% (728/882)
BRT – historical projected to current	0.87	81% (717/882)
GLM – historical projected to current	0.85	90% (798/882)

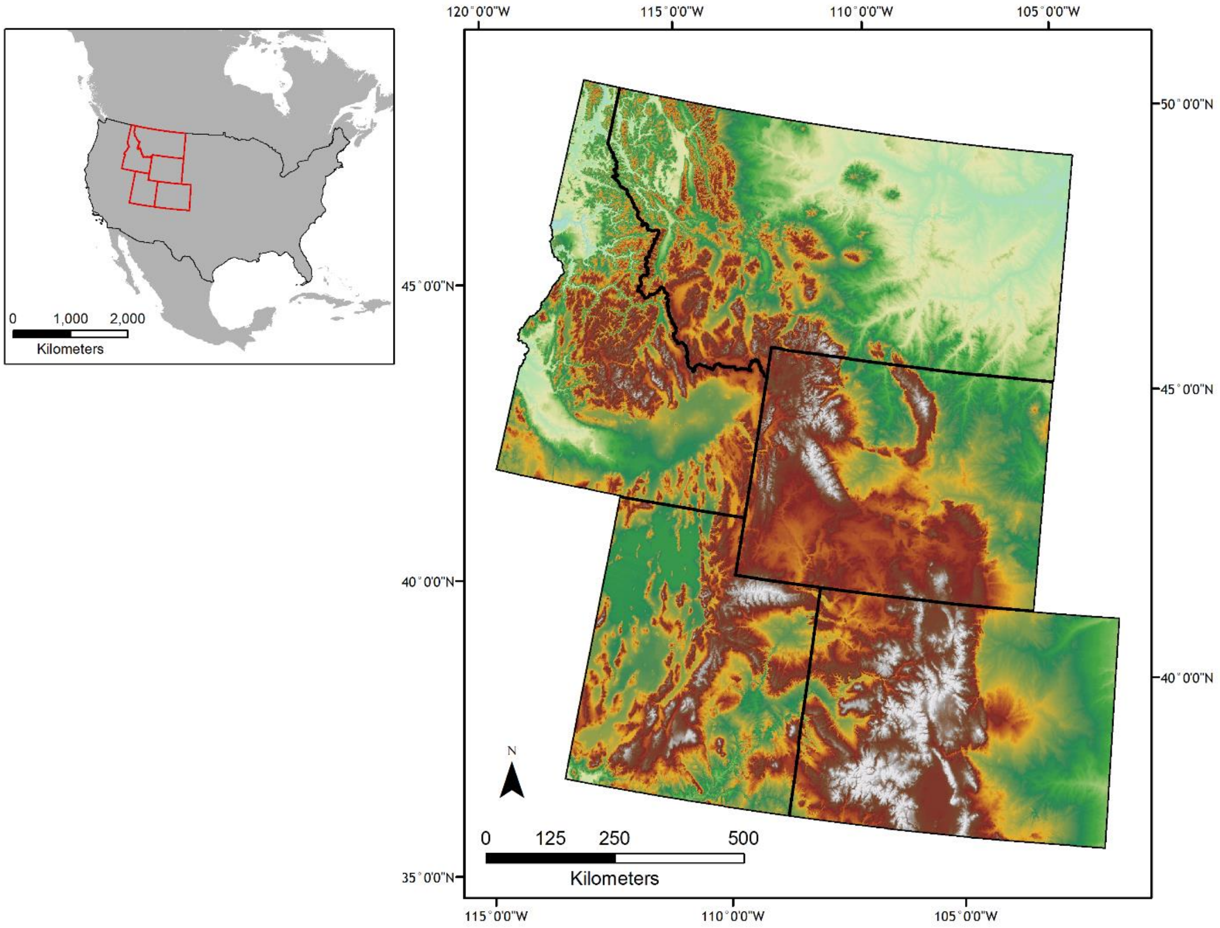

Figure 1: The five-state Rocky Mountain region comprising Colorado, Utah, Wyoming, Montana, and Idaho. The map shows topographic relief (1-km cells) across the region. Legend: elevations displayed as low elevation (140-1000 m, green), mid-elevation (1000-2500 m, orange/red), and high-elevation (> 2500 m, gray/white).

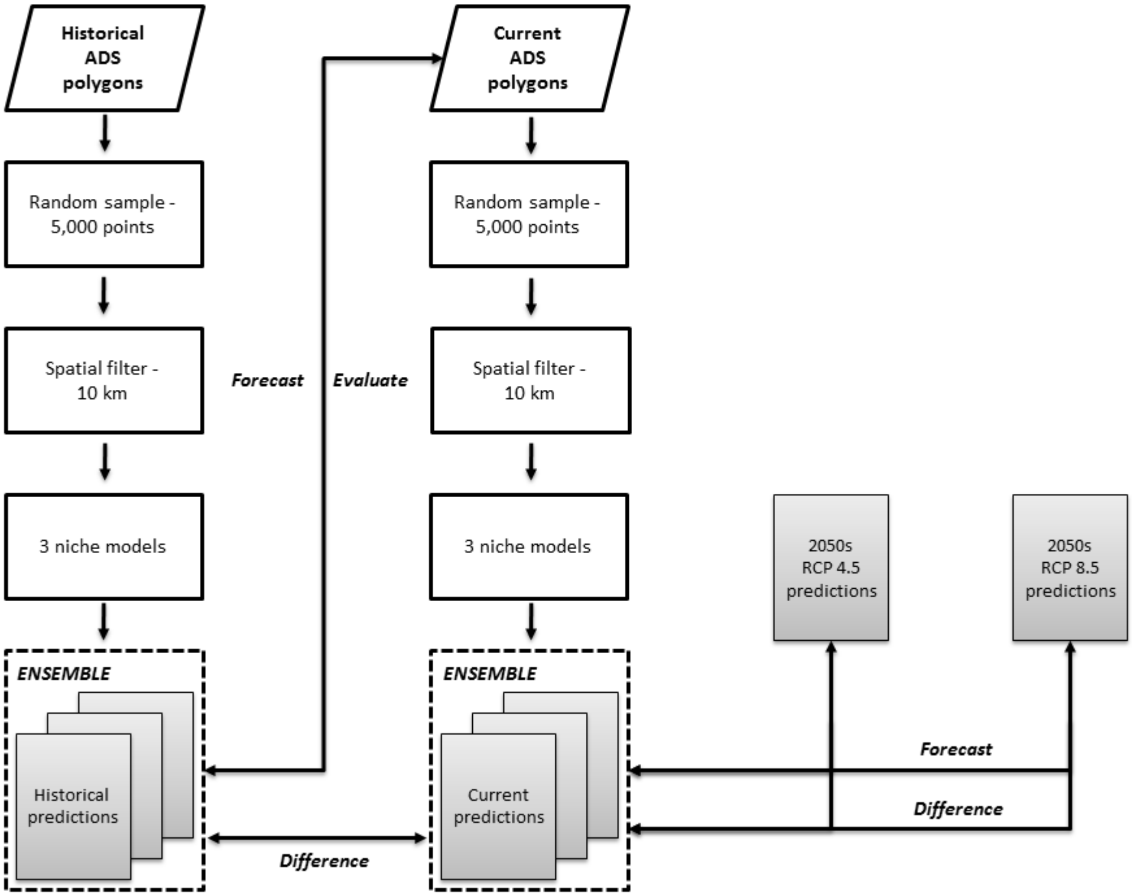

Figure 2: Schematic representation of the modeling design. See Methods for details; ADS is aerial detection survey.

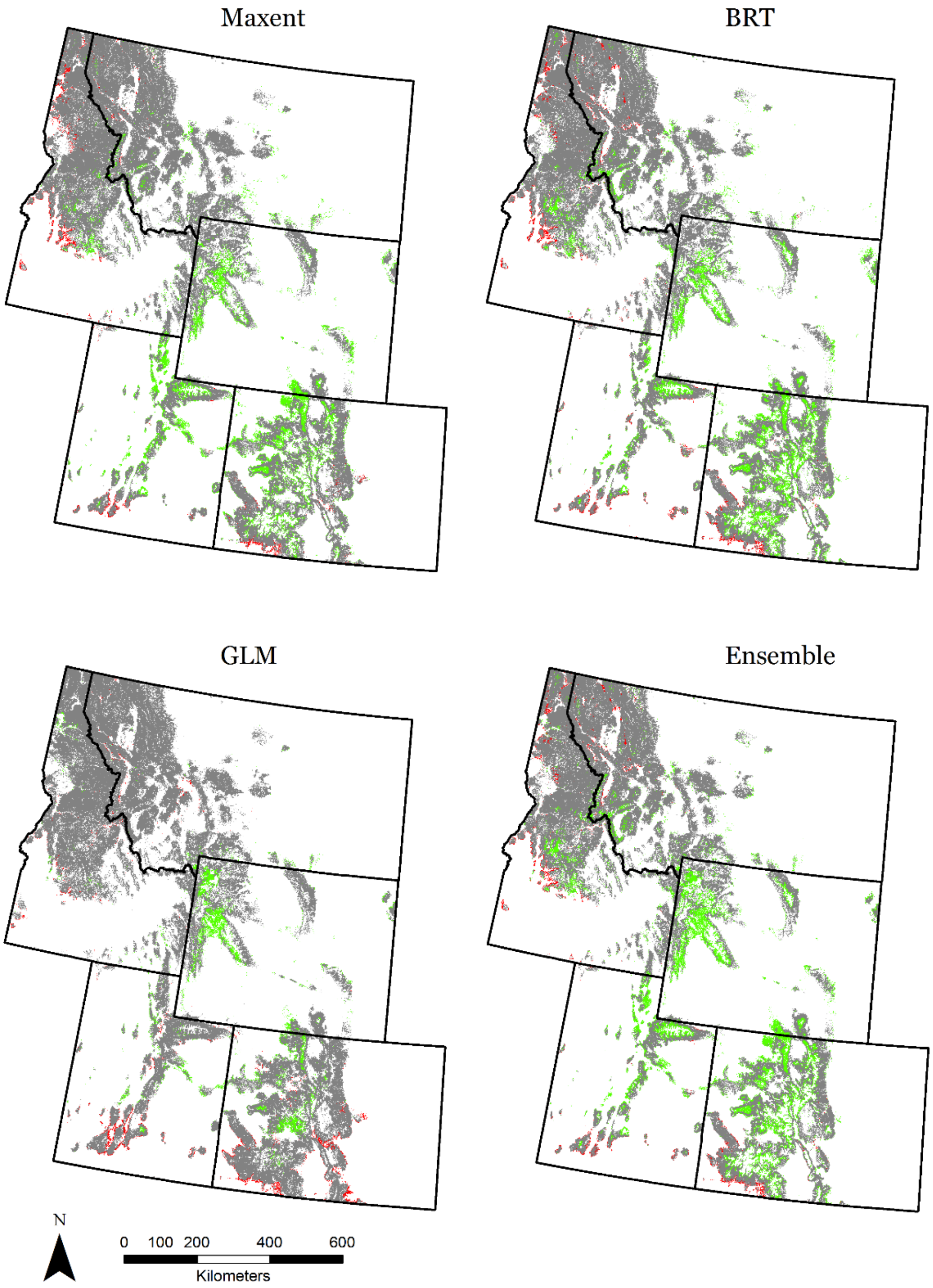

Figure 3: Estimated shift in climatically suitable areas between historical and current outbreaks. Legend: Suitable conditions in both outbreaks (gray), range expansion (green), range contraction (red).

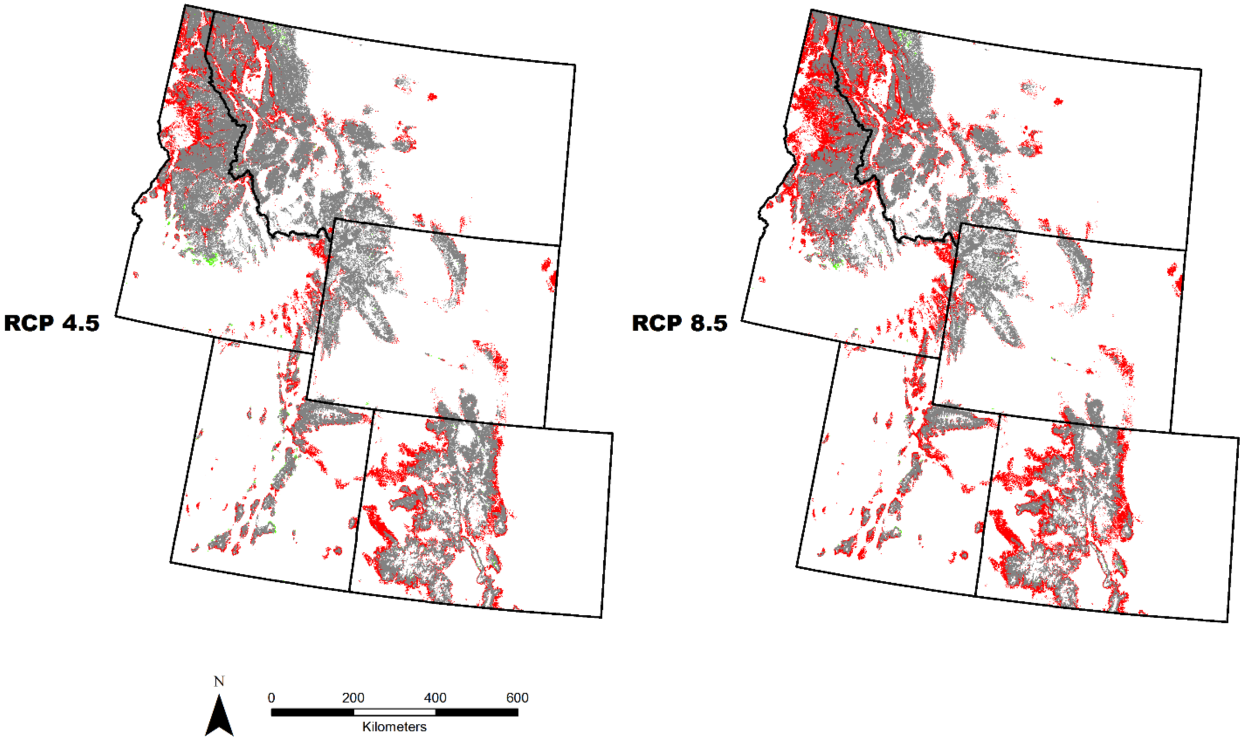

Figure 4: The ensemble models showing the shift in climatically suitable conditions under both the RCP 4.5 and RCP 8.5 future climate scenarios Legend: Climatically suitable habitat for current outbreak (gray), range expansion (green), range contraction (red).

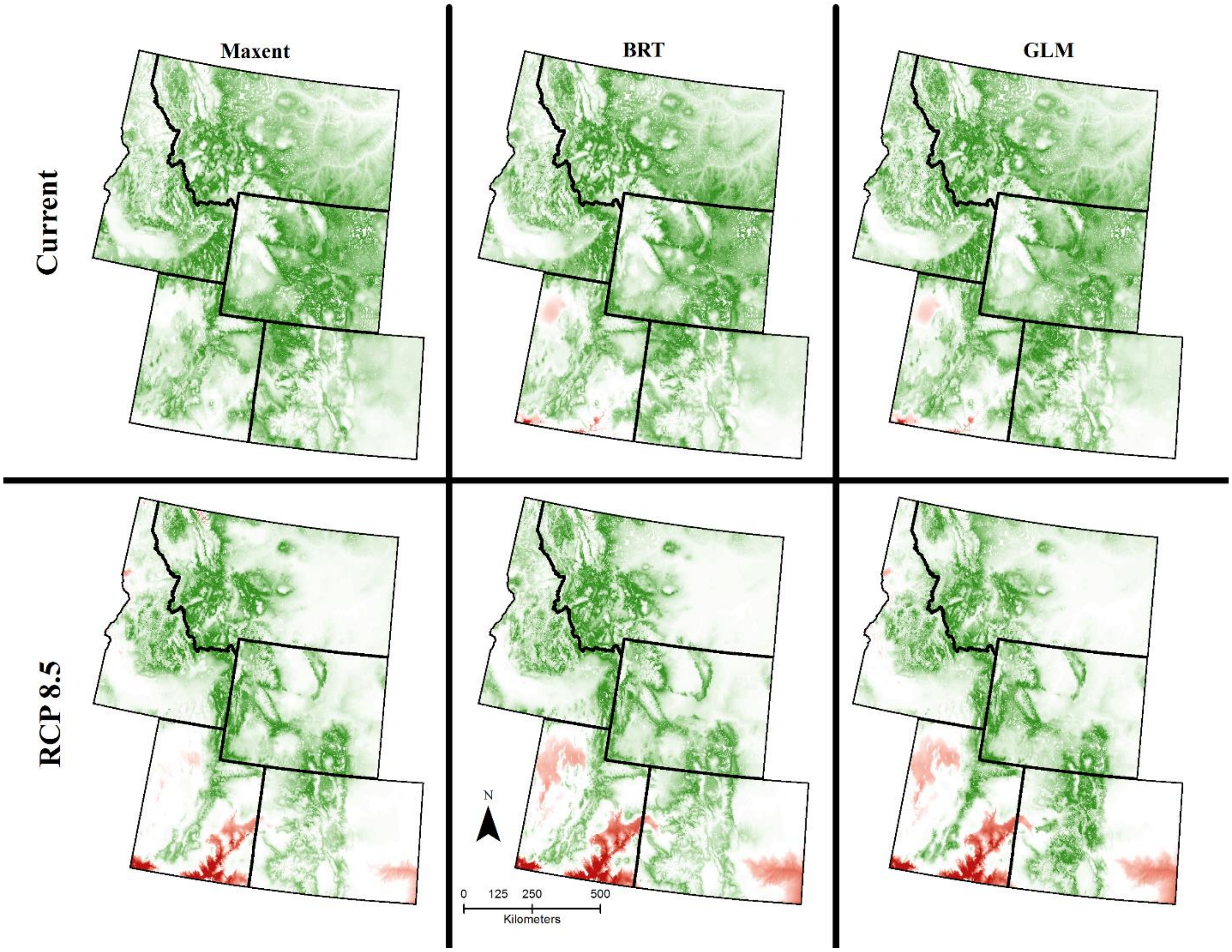

Figure 5: A comparison of predictor variables using Multivariate Environmental Similarity Surface (MESS) maps. The MESS calculation represents how similar a point is to a reference set of points. Negative values indicate novel environments where at least one variable has a value outside the range of environments found in the reference data. Sites with positive values indicate that the full range of environmental variables was found in the reference data; high positive values are fairly common and lower values represent a relatively unusual environment (Elith et al. 2010). Legend: green (positive), white (around zero), red (negative). Darker colors indicate more extreme values.

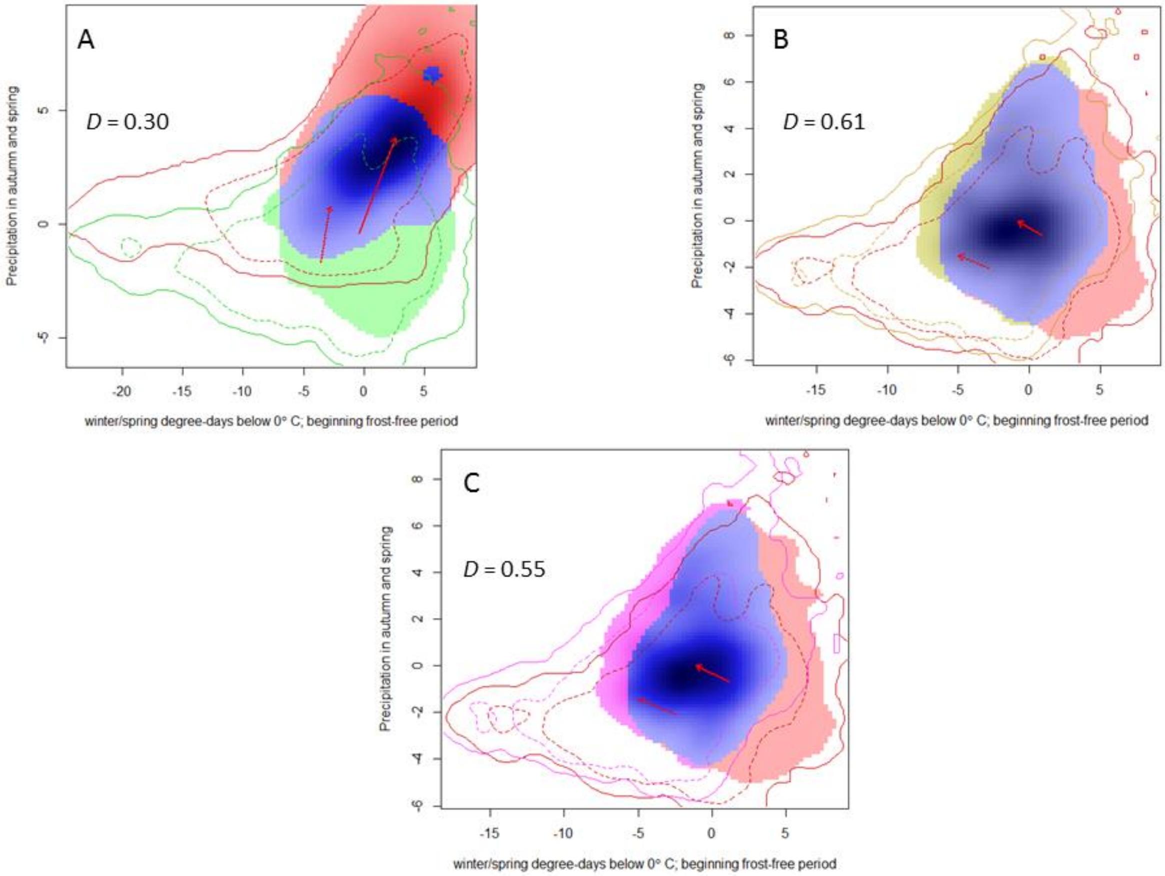

Figure 6: Principal component analysis of niche shift in environmental space for *D. ponderosae*. Blue shading represents overlap between periods. The solid and dashed contour lines illustrate, respectively, 100% and 50% of the available (background) environment. The solid arrows represent the shift of the niche for occupied sites and the dashed lines represent the shift across the full study area extent. Axes show the primary loadings of each principal component. (A) historical (green) and current (red); (B) current (red) and future under RCP 4.5 (gold); (C) current (red) and future under RCP 8.5 (purple).

REFERENCES

Amman, G. D. 1978. Biology, ecology, and causes of outbreaks of the mountain pine beetle in lodgepole pine forests. – In: Kibbee, D.L. et al. (eds), Theory and practice of mountain pine beetle management in lodgepole pine forests, 25-27, April 1978, Pullman, WA. Univ. of Idaho, Moscow, pp. 31–53.

Anderson, R. P. 2013. A framework for using niche models to estimate impacts of climate change on species distributions. - Ann. N. Y. Acad. Sci. 1297: 8–28.

Anderson, R. P. and Raza, A. 2010. The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: Preliminary tests with montane rodents (genus Nephelomys) in Venezuela. - J. Biogeogr. 37: 1378–1393.

Araújo, M. B. and New, M. 2007. Ensemble forecasting of species distributions. - Trends Ecol. Evol. 22: 42–7.

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

Araújo, M. B. et al. 2005. Validation of species – climate impact models under climate change. - Glob. Chang. Biol. 11: 1504–1513.

Assal, T. J. et al. 2014. Modeling a historical mountain pine beetle outbreak using Landsat MSS and multiple lines of evidence. - Remote Sens. Environ. 155: 275–288.

Aukema, B. H. et al. 2008. Movement of outbreak populations of mountain pine beetle: influences of spatiotemporal patterns and climate. – Ecography. 31: 348–358.

Austin, M. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. - Ecol. Modell. 157: 101–118.

Austin, M. P. 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. - Ecol. Modell. 200: 1–19.

Beaumont, L. J. et al. 2007. Where will species go? Incorporating new advances in climate modelling into projections of species distributions. - Glob. Chang. Biol. 13: 1368–1385.

Bentz, B. J. et al. 2010. Climate change and bark beetles of the western United States and Canada: Direct and indirect effects. - Bioscience 60: 602–613.

Beyer, H.L. 2012. Geospatial modelling environment (Version 0.7.3.0). http://www.spatialecology.com/gme.

Boria, R. A. et al. 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. - Ecol. Modell. 275: 73–77.

Broennimann, O. et al. 2012. Measuring ecological niche overlap from occurrence and spatial environmental data. - Glob. Ecol. Biogeogr. 21: 481–497.

Brown, J. L. 2014. SDMtoolbox: A python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. - Methods Ecol. Evol. 5: 694–700.

Campbell, E. 2007. Spatial distribution of mountain pine beetle outbreaks in relation to climate and stand characteristics: a dendroecological analysis. - J. Integr. Plant … 49: 168–178.

Carroll, A. L. et al. 2006. Impacts of climate change on range expansion by the mountain pine beetle. - Mt. Pine Beetle Initiat. 2006-14

Chapman, T. B. et al. 2012. Spatiotemporal patterns of mountain pine beetle activity in the southern Rocky Mountains. - Ecology 93: 2175–85.

Creeden, E. P. et al. 2014. Climate, weather, and recent mountain pine beetle outbreaks in the western United States. - For. Ecol. Manage. 312: 239–251.

Dale, V. H. et al. 2001. Climate change and forest disturbances. - Bioscience 51: 723.

de Oliveira, G. et al. 2014. Evaluating, partitioning, and mapping the spatial autocorrelation component in ecological niche modeling: a new approach based on environmentally equidistant records. - Ecography. 37: 001–011.

De'ath, G. 2007. Boosted trees for ecological modeling and prediction. - Ecology 88: 243–251.

Dobrowski, S. Z. et al. 2011. Modeling plant ranges over 75 years of climate change in California, USA: Temporal transferability and species traits. - Ecol. Monogr. 81: 241–257.

Elith, J. and Leathwick, J. R. 2009. Species distribution models: Ecological explanation and prediction across space and time. - Annu. Rev. Ecol. Evol. Syst. 40: 677–697.

Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. - Ecography. 29: 129–151.

Elith, J. et al. 2008. A working guide to boosted regression trees. - J. Anim. Ecol. 77: 802–13.

Elith, J. et al. 2010. The art of modelling range-shifting species. - Methods Ecol. Evol. 1: 330–342.

Evangelista, P. H. et al. 2011. Assessing forest vulnerability and the potential distribution of pine beetles under current and future climate scenarios in the Interior West of the US. - For. Ecol. Manage. 262: 307–316.

Fairweather, M.L., McMillin, J., Rogers, T., Conklin, D., and Fitzgibbon, B. 2006. Field guide to insects and diseases of Arizona and New Mexico forests. – USDA Forest Service Southwestern, Region MR-R3-16-3, p. 215.

Fieberg, J. 2007. Kernel density estimators of home range: smoothing and the autocorrelation red herring. - Ecology 88: 1059–1066.

Fielding, A. H. and Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. - Environ. Conserv. 24: 38–49.

Fitzpatrick, M. C. and Hargrove, W. W. 2009. The projection of species distribution models and the problem of non-analog climate. - Biodivers. Conserv. 18: 2255–2261.

Fry, J. A. et al. 2011. Completion of the 2006 national land cover database for the conterminous United States. - Photogramm. Eng. Remote Sensing 77: 858–566.

Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. - Ecol. Modell. 135: 147–186.

Guisan, A. and Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. - Ecol. Lett. 8: 993–1009.

Guisan, A. et al. 2007. What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? - Ecol. Monogr. 77: 615–630.

Heikkinen, R. K. et al. 2012. Does the interpolation accuracy of species distribution models come at the expense of transferability? - Ecography. 35: 276–288.

Hicke, J. A. et al. 2006. Changing temperatures influence suitability for modeled mountain pine beetle (*Dendroctonus ponderosae*) outbreaks in the western United States. - J. Geophys. Res. 1965: 1–23.

Homer, C. et al. 2007. Completion of the 2001 national land cover database for the conterminous United States. - Photogramm. Eng. Remote Sens. 73: 337–341.

Hubbard, R. M. et al. 2013. Changes in transpiration and foliage growth in lodgepole pine trees following mountain pine beetle attack and mechanical girdling. - For. Ecol. Manage. 289: 312–317.

IPCC 2014. Climate Change 2014: Synthesis Report. - In Core Writing Team, R.K. Pachauri and L.A. Meyer (eds), *Contribution of working groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. IPCC, Geneva, Switzerland, 151 pp.

Jewett, J. T. et al. 2011. Spatiotemporal relationships between climate and whitebark pine mortality in the Greater Yellowstone ecosystem. - For. Sci. 57: 320–335.

Jin, S. et al. 2013. A comprehensive change detection method for updating the national land cover database to circa 2011. - Remote Sens. Environ. 132: 159–175.

Johnson, E. W. and Ross, J. 2008. Quantifying error in aerial survey data. - Aust. For. 71: 216–222.

Khanum, R. et al. 2013. Predicting impacts of climate change on medicinal asclepiads of Pakistan using Maxent modeling. - Acta Oecologica 49: 23–31.

Kharouba, H. M. et al. 2009. Historically calibrated predictions of butterfly species' range shift using global change as a pseudo-experiment. - Ecology 90: 2213–2222.

Kumar, S. et al. 2014a. Assessing the potential for establishment of western cherry fruit fly using ecological niche modeling. - J. Econ. Entomol. 107: 1032–1044.

Kumar, S. et al. 2014b. Evaluating correlative and mechanistic niche models for assessing the risk of pest establishment. - Ecosphere 5: 86, *http://dx.doi.org/10.1890/ES14-00050.1*.

Lenoir, J. et al. 2008. A significant upward shift in plant species optimum elevation during the 20th century. - Science 320: 1768–71.

Lobo, J. M. et al. 2008. AUC: a misleading measure of the performance of predictive distribution models. - Glob. Ecol. Biogeogr. 17: 145–151.

Logan, J. A. and Bentz, B. J. 1999. Model analysis of mountain pine beetle (Coleoptera: Scolytidae) seasonality. - Environ. Entomol. 28: 924-934.

Logan, J. A. and Powell, J. A. 2001. Ghost forests, global warming, and the mountain pine beetle (Coleoptera : Scolytidae). - Am. Entomol. 47: 160–173.

Logan, J. A. et al. 2003. Assessing the impacts of global warming on forest pest dynamics. - Front. Ecol. Environ. 1: 130–137.

Lukas, J. and Gordon, E. 2015. Chapter 2—Colorado's climate: Past and future. In Gordon, E. and Ojima, D. (ed.), *Colorado Climate Change Vulnerability Study*. University of Colorado-Boulder Colorado State University.

Maestas, J. D. et al. 2011. Biodiversity and land-use change in the American mountain west. - Geogr. Rev. 91: 509–524.

McCambridge, W. 1971. Temperature limits of flight of the mountain pine beetle, *Dendroctonus ponderosae*. - Ann. Entomol. Soc. Am. 64: 534–535.

McConnell, T. et al. 2000. A guide to conducting aerial sketchmapping surveys. – FHTET 00-01. U.S. Department of Agriculture, Forest Service, Fort Collins, CO.

McCullagh, P. and Nelder, J.A. 1989. Generalized linear models. – Chapman and Hall.

Merow, C. et al. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. – Ecography. 36: 1058–1069.

Monahan, W. B. et al. 2013. Forecasting distributional responses of limber pine to climate change at management-relevant scales in rocky mountain national park. - PLoS One 8: e83163.

Morisette, J. T. et al. 2013. VisTrails SAHM: visualization and workflow management for species habitat modeling. - Ecography. 36: 129–135.

Moss, R. H. et al. 2010. The next generation of scenarios for climate change research and assessment. - Nature 463: 747–56.

Muscarella, R. et al. 2014. ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. - Methods Ecol. Evol. 5: 1–8.

Negrón, J. F. and Fettig, C. J. 2014. Mountain pine beetle, a major disturbance agent in US western coniferous forests: A Synthesis of the state of knowledge. - For. Sci. 60: 409–413.

Parmesan, C. 2006. Ecological and evolutionary responses to recent climate change. - Annu. Rev. Ecol. Evol. Syst. 37: 637–669.

Pearman, P. B. et al. 2008. Prediction of plant species distributions across six millennia. - Ecol. Lett. 11: 357–369.

Pearson, R. G. and Dawson, T. P. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? - Glob. Ecol. Biogeogr. 12: 361–371.

Pearson, R. G. et al. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. - J. Biogeogr. 34: 102–117.

Peterson, N. et al. 2011. Ecological niches and geographic distributions. – Princeton Univ. Press.

Phillips, S. J. et al. 2006. Maximum entropy modeling of species geographic distributions. - Ecol. Modell. 190: 231–259.

Phillips, S. et al. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. - Ecol. Appl. 19: 181–197.

Radosavljevic, A. and Anderson, R. P. 2014. Making better Maxent models of species distributions: complexity, overfitting and evaluation. - J. Biogeogr. 41: 629–643.

Raffa, K. F. et al. 2008. Cross-scale drivers of natural disturbances prone to anthropogenic amplification: The dynamics of bark beetle eruptions. - Bioscience 58: 501.

Régnière, J. and Bentz, B. J. 2007. Modeling cold tolerance in the mountain pine beetle, *Dendroctonus ponderosae*. - J. Insect Physiol. 53: 559–72.

Safranyik, L. 1978. Effects of climate and weather on mountain pine beetle populations. – In: Kibbee, D.L. et al. (eds), Theory and practice of mountain pine beetle management in lodgepole pine forests, 25-27, April 1978, Pullman, WA. Univ. of Idaho, Moscow, pp. 77–84.

Safranyik, L. and Carroll, A. L. 2006. The biology and epidemiology of the mountain pine beetle in lodgepole pine forests. - In: Safranyik, L. and Wilson, W. R. (eds), The Mountain pine beetle: A synthesis of biology, management, and impacts on lodgepole pine. Pacific Forestry Centre, pp. 3–66.

Safranyik, L. et al. 1975. An interpretation of the interaction between lodgepole pine, the mountain pine beetle and its associated blue-stain fungi in western Canada. – In: Baumgartner, D. M. (ed.), Management of lodgepole pine ecosystems. Washington State Univ. Cooperative Extension Service, Pullman, pp. 406 – 428.

Safranyik, L. et al. 2010. Potential for range expansion of mountain pine beetle into the boreal forest of North America. - Can. Entomol. 142: 415–442.

Sambaraju, K. R. et al. 2012. Climate change could alter the distribution of mountain pine beetle outbreaks in western Canada. - Ecography. 35: 211–223.

Schroeder B. 2006. ROC/AUC-calculation. - http://brandenburg.geoecology.unipotsdam.de/users/schroeder/download.html.

Seager, R. et al. 2007. Model projections of an imminent transition to a more arid climate in southwestern North America. - Science 316: 1181–1184.

Shcheglovitova, M. and Anderson, R. P. 2013. Estimating optimal complexity for ecological niche models: A jackknife approach for species with small sample sizes. - Ecol. Modell. 269: 9–17.

Stohlgren, T. J. et al. 2010. Ensemble habitat mapping of invasive plant species. - Risk Anal. 30: 224–35.

Taylor, S. and Carroll, A. 2003. Disturbance, forest age, and mountain pine beetle outbreak dynamics in BC: A historical perspective. - Mt. Pine Beetle Symp. Challenges Solut.: 41–51.

Taylor, K. E. et al. 2012. An overview of CMIP5 and the experiment design. - Bull. Am. Meteorol. Soc. 93: 485–498.

Thuiller, W. et al. 2008. Predicting global change impacts on plant species' distributions: Future challenges. - Perspect. Plant Ecol. Evol. Syst. 9: 137–152.

U.S. Forest Service (USFS) 2011. Western bark beetle strategy: Human safety, recovery and resiliency. – USDA Forest Service.

USFS 2014. Insect and disease detection survey database (IDS). Digital Data. - Forest Health Technology Enterprise Team. http://foresthealth.fs.usda.gov/ids.

van Mantgem, P. J. et al. 2009. Widespread increase of tree mortality rates in the western United States. - Science 323: 521–4.

van Vuuren, D. P. et al. 2011. The representative concentration pathways: An overview. - Clim. Change 109: 5–31.

Veloz, S. D. 2009. Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. - J. Biogeogr. 36: 2290–2299.

Wang, T. et al. 2012. ClimateWNA—High-resolution spatial climate data for western North America. - J. Appl. Meteorol. Climatol. 51: 16–29.

Warren, D. L. et al. 2008. Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. - Evolution, 62: 2868–2883.

Wenger, S. J. and Olden, J. D. 2012. Assessing transferability of ecological models: An underappreciated aspect of statistical validation. - Methods Ecol. Evol. 3: 260–267.

West, A. M. et al. 2015. Using high-resolution future climate scenarios to forecast *Bromus tectorum* invasion in Rocky Mountain National Park. - PLoS One 10: e0117893.

Westerling, A. L. et al. 2006. Warming and earlier spring increase western U.S. forest wildfire activity. - Science 313: 940–943.

Wiens, J. A. et al. 2009. Niches, models, and climate change: assessing the assumptions and uncertainties. - Proc. Natl. Acad. Sci. U. S. A. 106: 19729–19736.

Williams, J. W. et al. 2007. Projected distributions of novel and disappearing climates by 2100 AD. - Proc. Natl. Acad. Sci. U. S. A. 104: 5738–5742.

CHAPTER 2
CONCLUSIONS AND REFLECTIONS

To better understand the future of North American forests under changing climatic regimes, it is imperative to understand the conditions that have supported large mountain pine beetle outbreaks over the past 50 years. This research used interesting new approaches to species distribution modeling to advance the understanding of how climate shaped the distribution of the mountain pine beetle under historical and current conditions and estimated distributions under future climate scenarios. My results indicate that there has been a significant expansion of suitable environments for the MPB over the past half-century in both geographic and environmental space; however, projected warming may reduce this suitable habitat under future climate scenarios. Furthermore, my findings show that suitable habitats have shifted upwards in elevation and may continue to do so in the future. This shift threatens sensitive high-elevation ecosystems, like those dominated by whitebark pine (Jewett et al. 2011), but may also reflect the destabilization of currently suitable habitats at lower elevations. My results confirm that climate change has driven a range expansion of MPB and corroborates past research on the effects of climate on the spatial distribution of MPB outbreaks.

The three SDM used in this analysis provide fairly reliable estimates of MPB’s response to climate change. While studies utilizing SDM should always be aware of the assumptions and limitations of the models (Pearson and Dawson 2003, Araújo and Peterson 2012), SDMs can be an effective and reliable tool in predicting change across temporal domains. When choosing a model, simpler algorithms, like the GLM, may provide general predictions that better project across temporal domains. Confidence in model predictions can be bolstered by using MESS maps that show areas of extrapolation in model projections (Elith et al. 2010). These features are useful tools that are available in both Maxent and SAHM. In this study, the MESS maps indicate that future conditions in the study area, particularly in the forested environments that host MPB, may not represent novel conditions due to the heterogeneous landscape of the study area. The MESS analysis helps to reduce uncertainty in SDM projections, though it cannot account for the uncertainty inherent in the climate forecasts themselves.

Species distribution modeling is a rapidly growing and advancing field that has become one of the more prominent sub-fields in ecological research over the past 15 years (Pearson and Dawson 2003, Elith et al. 2006, Elith et al. 2010, Peterson et al. 2011, Araújo and Peterson 2012). The three models I used in this research are common, high-performing algorithms, but they are commonly misused or are not parameterized to optimize results (Elith et al. 2008, Merow et al. 2013, Radosavljevic and Anderson 2014). In addition to the insight on MPB and the effects of climate change in the Rocky Mountain region, my research was intended, in part, to demonstrate the current best practices associated with these models. Proper parameterization is critical to correctly fit models to the species of interest, sample size, and the environmental data. By utilizing tools such as the “ENMeval” package (Muscarella et al. 2014) and various functions in SAHM (Morisette et al. 2013), model performance can be optimized by moving beyond the default settings.

Reflections

Overall, I have greatly enjoyed my time at NREL and GDPE. I have been repeatedly impressed with the breadth of expertise and the diversity of research within both entities, and am grateful to have had the opportunity to study here. My time here has not always been easy, but I have grown a great deal from my experiences and believe they have made me a better scientist. This final section of my thesis contains my reflections on growing as a modeler, my graduate experience, and suggestions for incoming graduate students.

On Modeling

Graduate school is littered with “aha moments.” Sometimes these moments reflect relatively small triumphs—cracking an obstinate chunk of code or finding the elusive words to end a paper—but sometimes they provide a moment of clarity that helps elucidate a key concept or question. For me, an important “aha” moment helped me to better understand the “why” of ecological models. One of my favorite courses in the Graduate Degree Program in Ecology (GDPE) was taught by Dr. Tom Hobbs, Models for Ecological Data. Early in the course, Dr. Hobbs came to class with a fly fishing rod, a recording of “Madame Butterfly” by Giacomo Puccini, and a print of “Starry Night” by Vincent van Gogh; somehow these items all related to Bayesian statistics and modeling. These seemingly disparate items all had one characteristic in common: they represented abstractions, or simplifications, of reality. He went on to explain that models of ecological phenomena are also just abstractions of complex and dynamic systems. George E.P. Box (1987) is widely credited with saying, “Essentially, all models are wrong, but some are useful,” and Dr. Hobbs helped me to truly understand what this means.

This lesson in abstraction has played a part in defining my approach to SDMs and has aided in the interpretation of modeling research. Models can be incredibly useful tools for deriving important relationships between a species and its environment, and can elucidate general trends across time or space. However, there is a tendency among modelers to over-emphasize results, especially when projecting future responses to novel conditions. This has resulted in two broad opinions on the utility of SDMs: those who believe in their usefulness and those who do not. I have experienced this dichotomy in my own research, and I believe that my awareness of it has made me a better modeler and scientist.

There are two particular papers that helped clarify my thoughts on the proper usage of SDMs: Araújo and Peterson (2012), which eloquently summarized the uses and misuses of SDM, and an earlier paper by Pearson and Dawson (2003). Combined with my own experience, these studies have greatly informed my approach to correlative modeling. First, scale matters! The scale of the variables used and the extent of the study region should be defined by the processes being examined. For example, assessment of climatic interactions should be defined at regional to continental scales (> 200 km) where large-scale processes are more likely to influence distribution. Variables like land-use and dispersal play a far greater role in shaping a species distribution at smaller scales (Pearson and Dawson 2003). The details of the modeling process (e.g. variables used, grain size, extent) should be scaled based on the processes examined.

Secondly, it is critical to investigate any biases and assumptions in the modeling process. Sampling bias can misinform models, so it is important to thoroughly understand and vet the data used in the models. There are many ways to address sampling bias in the modeling process (Anderson and Gonzalez 2011, Barbet-Massin et al 2012, Boria et al. 2014), and these should be implemented to produce the best possible models. Some examples of how to do this include spatial filtering of the occurrence data and restricting the selection of background pseudo-absences to general areas of use for the species. Other assumptions and uncertainties, like species-environment equilibrium (Wiens et al. 2009), are more difficult to address but should be accounted for in the discussion and interpretation of the results. Uncertainty is inherent in most data sources, particularly models of future climatic conditions, and this should be explicitly discussed.

Uncertainty is largely unavoidable in ecological modeling. Some approaches, like Bayesian models, quantify uncertainty, but this is less frequent with correlative models. This uncertainty does not invalidate models, but it should guide the interpretation and application of correlative models. Outputs from these models should be treated as distributional hypotheses (Lobo et al. 2008), and particularly at larger scales, are more useful in defining general trends (e.g. upwards range expansion) than predicting locations of occurrence. Though many modelers use SDMs for prediction, I am of the opinion that this is generally an inappropriate usage. When prediction is the goal, it should be used in conjunction with field sampling and monitoring to validate results. There is danger in wholeheartedly believing in model predictions, but they are nevertheless useful tools for identifying important relationships and possible future trends.

I am a firm believer in the utility of models and have enjoyed coming to understand their application and function. As I move forward in my career, I will continue using many of the models in my research, but I hope to also broaden my future usage by incorporating more field validation and empirical analysis.

On graduate school

Coming into GDPE, I was primarily interested in the quantitative research tools (e.g. GIS, remote sensing, and statistical modeling) used in ecological analysis, but over the past three years, I have grown particularly interested in how climate change shapes large-scale landscape patterns. I am pleased that my research incorporated landscape-level change, disturbance, and the impacts of climate change, and I hope I have contributed something worthwhile to the field. With every passing year, the implications of climate change become direr, and ecologists are increasingly asked to assess the effects of this change and the resulting impacts on ecosystems, now and in the future. I believe my research has provided me with the skills and knowledge needed to address the impacts of climate change, and a good foundation on which to continue building my career. Overall, I am excited by the career possibilities in front of me.

Of course, there were many challenges to graduate school, and I would like to spend a little time imparting my own lessons-learned to ease the process for future students. What follows are my thoughts on how to best tackle grad school. It is, by no means, the only path to success, but I hope it is helpful to any readers of this document. When I entered graduate school, I was primarily interested in quantitative research tools more so than specific ecological systems or questions, so I spent my first year focused on learning these tools and less time familiarizing myself with foundational ecological principles. As a result, the tools guided the development of my research in its early stages instead of interesting ecological questions. I ultimately came to understand that the question should guide the methods, and working backwards from a broad ecological question—e.g. how does climate change drive MPB outbreaks?—will open up more research possibilities than trying to fit a tool to a problem. Also, from a practical perspective, taking courses on quantitative methods at a more advanced stage of one’s degree allows for the incorporation of class projects as elements of the thesis.

It is easy to get caught up in the reading, writing, and coding of grad school, but it is essential to take the time to just be in the environment you study. Do not forget to spend time in the study system; if that is not possible, at least spend time outside in natural settings. Further, take time to just sit and observe with all your senses. These periods of reflection and observation are essential for grasping how a system functions and will ultimately yield a deeper understanding of any results from modeling or geospatial analyses. Observation is at the root of the scientific method, so take the time to truly observe what is happening in your system.

It is vital to find an academic community while in graduate school. In my experience, it is incredibly important to find other scientists working in the same system and addressing related questions or using the same tools. This community will challenge you to improve your research and can be a significant support during the challenging periods that are inherent to graduate school. Having a friend to read a draft or help troubleshoot a technical difficulty can make a big difference in both the quality and enjoyment of one’s work. Conversely, graduate school is a wonderful time to intellectually explore a wide variety of research, so do not restrict yourself only to topics that relate to your research. Reading broadly (and attending lectures) will help provide perspective on your work and that of your colleagues. Occasionally, methods used in other fields may also apply to a problem in ecology.

I have devoted a great deal of time and effort to this document and believe I have positively contributed to the understanding of the impacts of climate change on the mountain pine beetle. Along the way, I learned some powerful and interesting tools that will benefit me throughout my career. I reveled in the intellectual stimulation of graduate school and leave with many close friends and terrific memories.

REFERENCES

Anderson, R. P. and Gonzalez, I. 2011. Species-specific tuning increases robustness to sampling bias in models of species distributions: An implementation with Maxent. - Ecol. Modell. 222: 2796–2811.

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

Barbet-Massin, M. et al. 2012. Selecting pseudo-absences for species distribution models: How, where and how many? - Methods Ecol. Evol. 3: 327–338.

Boria, R. A. et al. 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. - Ecol. Modell. 275: 73–77.

Box, G. E., & Draper, N. R. 1987. Empirical model-building and response surfaces. - John Wiley & Sons.

Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. - Ecography. 29: 129–151.

Elith, J. et al. 2008. A working guide to boosted regression trees. - J. Anim. Ecol. 77: 802–13.

Elith, J. et al. 2010. The art of modelling range-shifting species. - Methods Ecol. Evol. 1: 330–342.

Jewett, J. T. et al. 2011. Spatiotemporal Relationships between Climate and Whitebark Pine Mortality in the Greater Yellowstone Ecosystem. - For. Sci. 57: 320–335.

Lobo, J. M. et al. 2008. AUC: a misleading measure of the performance of predictive distribution models. - Glob. Ecol. Biogeogr. 17: 145–151.

Merow, C. et al. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. – Ecography. 36: 1058–1069.

Morisette, J. T. et al. 2013. VisTrails SAHM: visualization and workflow management for species habitat modeling. - Ecography. 36: 129–135.

Muscarella, R. et al. 2014. ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. - Methods Ecol. Evol. 5: 1–8.

Pearson, R. G. and Dawson, T. P. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? - Glob. Ecol. Biogeogr. 12: 361–371.

Peterson, N. et al. 2011. Ecological niches and geographic distributions. – Princeton Univ. Press.

Radosavljevic, A. and Anderson, R. P. 2014. Making better Maxent models of species distributions: complexity, overfitting and evaluation (M Araújo, Ed.). - J. Biogeogr. 41: 629–643.

Wiens, J. A. et al. 2009. Niches, models, and climate change: assessing the assumptions and uncertainties. - Proc. Natl. Acad. Sci. U. S. A. 106: 19729–19736.

APPENDIX 1 – STUDY AREA

Recent bark beetle infestations began in the late 1990s and have severely impacted the coniferous forests of the western U.S. Five states in particular have experienced, and continue to experience, extensive outbreaks (USFS 2011, 2014a): Colorado, Utah, Wyoming, Montana, and Idaho. These five states largely comprise the U.S. Rocky Mountain region and define this project's study area (Figure 1). Climate, topography, and vegetative cover vary across the extent. Temperature is intricately linked to topography and the region is prone to extended periods of drought. The montane and subalpine forests of the Rocky Mountains are characterized by the dominance of coniferous species including ponderosa pine, Douglas-fir (*Pseudotsuga menziesii*), lodgepole pine, whitebark pine, subalpine fir (*Abies lasiocarpa*), and Engelmann spruce (*Picea engelmannii*) (Evangelista et al. 2011). The eastern margins of the study area include the shortgrass prairies of the Great Plains and semi-arid shrubs, while pinyon pine (*Pinus edulis*), and juniper (*Juniperus spp.*) woodlands dominate the arid southwest portion of the study area (Evangelista et al. 2011).

APPENDIX 2 – CLIMATE DATA

The ClimateWNA software provides over 20,000 surfaces of monthly, seasonal, and annual climate data for point locations and time series across western North America. ClimateWNA offers a suite of directly calculated and derived monthly, seasonal, and annual climatic variables based on the 30-year normal for the chosen time period. This software uses the delta approach where historical data (CRU TS 2.1) and future projections are expressed and interpolated as a difference from the baseline reference period (PRISM and ANUSPLIN grids). These interpolations are tested against local weather station data. Climate data are downscaled based a digital elevation model (DEM) using partial derivative functions of temperature change along elevational gradients (Wang et al. 2012). Topographic variables were generated from a 1-km DEM acquired from ClimateWNA and clipped to the study area; all processing was done in ArcGIS (ESRI v10.2).

ClimateWNA includes a suite of 15 general circulation models (GCMs) selected from phase five of the Coupled Model Intercomparison Project (CMIP5) multi-model data set that corresponds with the Fifth Assessment Report from the Intergovernmental Panel on Climate Change (IPCC) (Taylor et al. 2012, Wang et al. 2012). The RCP framework was developed after the IPCC Fourth Assessment Report (IPCC 2007) and uses scenarios based on future radiative forcings (Moss et al. 2010). For this analysis I used a global mean of the 15 GCMs for my future climate scenarios for both RCP 4.5 and RCP 8.5.

Table 5: Initial set of climatic and topographic variables considered for the model before correlation analysis. All variables acquired from ClimateWNA; for a full description of the variables see Wang et al. (2012).

Variable	Description
Directly calculated annual variables	
MAT	Mean annual temperature (°C)
MWMT	Mean warmest month temperature (°C)
MCMT	Mean coldest month temperature (°C)
MAP	Mean annual precipitation (mm)
MSP	Mean summer (May to Sept) precipitation (mm)
Derived annual variables	
DD>5	Degree-days above 5°C, growing degree-days
EMT	Extreme minimum temperature over 30 years.
CMD	Hargreaves climatic moisture deficit, a measure of potential evapotranspiration
bFFP	The Julian date on which frost-free period begins
eFFP	The Julian date on which frost-free period ends
NFFD	The number of frost-free days
PAS	Precipitation as snow (mm) between August in previous year and July in current year
Directly calculated seasonal variables	
Tave_sp	Spring mean temperature (°C)
Tave_sm	Summer mean temperature (°C)
Tave_at	Autumn mean temperature (°C)
Tmax_sp	Spring mean maximum temperature (°C)
Tmax_sm	Summer mean maximum temperature (°C)
Tmax_at	Autumn mean maximum temperature (°C)
Tmin_wt	Winter mean minimum temperature (°C)
Tmin_sp	Spring minimum temperature
PPT_sp	Spring precipitation (mm)
PPT_sm	Summer precipitation (mm)
PPT_at	Autumn precipitation (mm)
Derived seasonal variables	
DD_0_wt	Winter degree-days below 0°C
DD_0_sp	Spring degree-days below 0°C
DD5_sp	Spring degree-days above 5°C
DD5_at	Autumn degree-days above 5°C

DD18_sp	Spring (Mar-May) degree-days above 18°C
DD18_sm	Summer (June-August) degree-days above 18°C
DD18_at	Autumn (Sep-Nov) degree-days above 18°C
CMD_sm	Summer Hargreaves climatic moisture deficit (mm)
CMD_at	Autumn Hargreaves climatic moisture deficit (mm)
Monthly variables	
Tmax	Maximum mean temperatures (°C) for July, August, September, October, and November
DD18	Degree-days above 18°C for July, August, September, October, and November
Topographic variables	
Elevation	Digital elevation model (DEM) at 1 km resolution
Slope	Maximum change in elevation between each cell and its eight neighbors. Calculated from the DEM in ArcGIS
Aspect	Downslope direction of a grid cell

APPENDIX 3 – CORRELATIVE NICHE MODELS

MaxEnt

MaxEnt is a maximum-entropy approach that finds a probability distribution defined over the study area that satisfies a set of constraints based on the occurrence data; these constraints rely upon environmental variables that serve as predictor variables (Phillips et al. 2006, Guisan et al., 2007). MaxEnt is a particularly powerful tool for handling presence-only data. In addition to being easy to use (Merow et al., 2013), MaxEnt generally outperforms other modeling methods on predictive accuracy (Elith et al., 2006; Austin, 2007; Guisan et al., 2007; Stohlgren et al., 2010; Merow et al., 2013).

A distinct advantage of MaxEnt is its approachability and ease of use; the graphical user interface allows the user to easily import data and generate models using the default settings. This can lead to models that are lacking in quality and predictive power, and a number of recent studies have underscored the importance of carefully calibrating the MaxEnt model (Merow et al. 2013; Radosavljevic and Anderson 2013; Shcheglovitova and Anderson 2014). Two settings in particular, 'features class' and the 'regularization multiplier', can have drastic effects on model fitting. Features are simple functions derived from environmental variables, mathematical transformations that constrain the predictor variables (Phillips and Dudík 2008, Merow et al. 2013). MaxEnt can be run with any single class or various combinations of the five feature classes: linear (L), quadratic (Q), product (P), threshold (T), and hinge (H). The choice of feature class(es) to include is related to the sample size of the dataset; the "auto features" selection includes all five feature and has been shown to be the optimal combination for sample sizes greater than 80 (Phillips and Dudík, 2008). Regularization is a limit on model complexity in MaxEnt, a penalty for each term included in the model and for higher weights given to a term (Radosavljevic and Anderson 2013). MaxEnt uses a regularization multiplier, a coefficient that is applied to the parameter of each feature class that changes the overall level of regularization but not the parameters individually (Radosavljevic and Anderson 2013).

To calibrate the MaxEnt models for the mountain pine beetle, I experimentally tuned using the "ENMeval" package in R (Muscarella et al., 2014). The ENMeval package builds a series of candidate models in MaxEnt with a variety of settings and then provides multiple evaluation metrics to help select the optimal model (Muscarella et al., 2014). The default settings in MaxEnt (tested by Phillips and Dudík, 2008) include all feature classes (LQHTP) and a regularization multiplier of 1.0. ENMeval was set up to test four feature class combinations (LQH, LQHP, LQHT, and LQHPT) and seven regularization multipliers (1.0, 1.5, 2.0, 2.5, 3.0, 3.5, and 4.0) for a total of 28 model combinations for both the historic and current models. Four evaluation metrics were assessed with ENMeval: the Akaike's Information Criterion corrected for small sample sizes (AICc), area under the receiver operating curve (AUC), the difference between training and test AUC values (AUCDIFF), and the 'minimum training presence' omission rate (ORMIN) (Muscarella et al. 2014). The ENMeval results indicate that the optimal feature class combination for both the historical and current models was the "all features" default; however, the AUCDIFF values show that the models overfit at low regularization multipliers (Figures 7, 8).

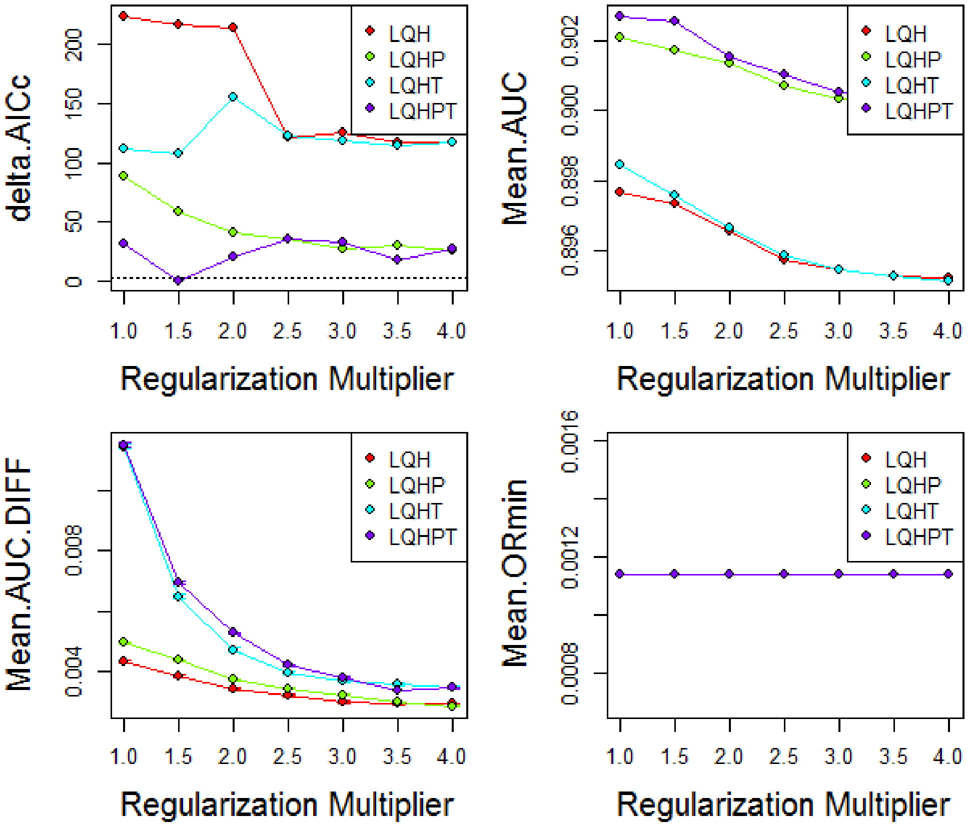

Figure 7: Evaluation metrics generated by ENMeval for the historic dataset.

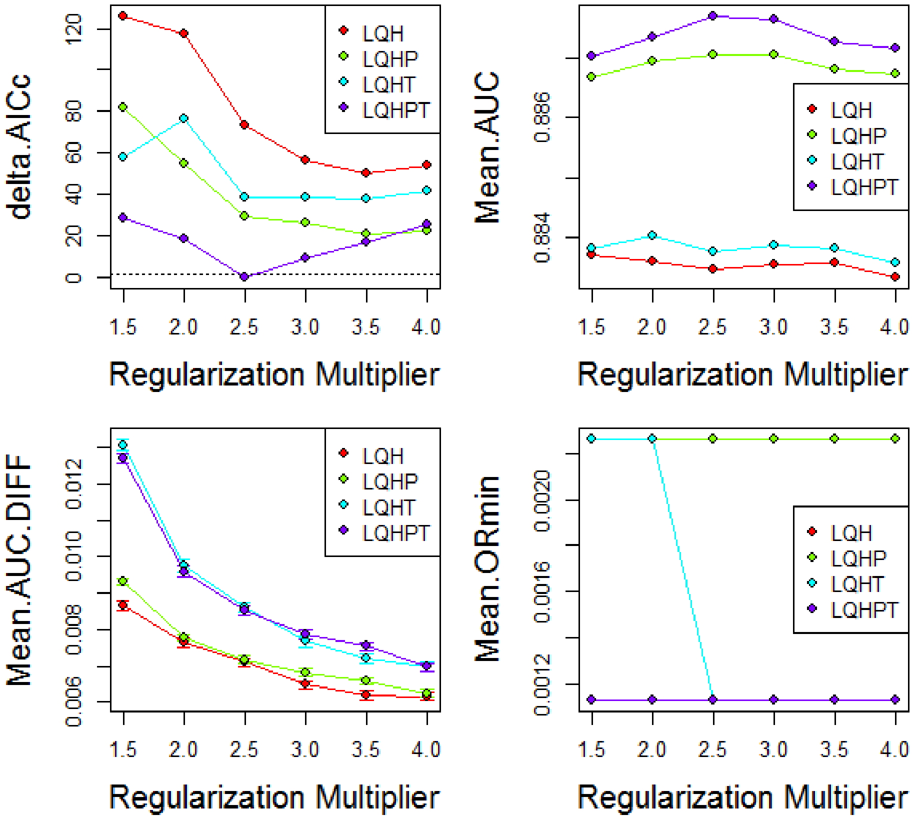

Figure 8: Evaluation metrics generated by ENMeval for the current data set.

Boosted Regression Trees

Boosted regression trees handle a variety of predictor variable types, have no need for prior data transformation, can fit complex nonlinear relationships, and handle interaction effects between predictors (Elith et al., 2008). The BRT begins with a single decision tree then adds a tree that best explains the error from the first tree and so on until the model is completed (Morisette et al., 2013). Compared to other modeling methods, BRT have performed well and are generally recognized as having strong predictive success similar to MaxEnt (Elith et al., 2006; Austin, 2007; Guisan et al., 2007). The BRT modeling procedure was run through the VisTrails:SAHM framework (Morisette et al., 2013).

The BRT implementation in SAHM incorporates an algorithm for tuning the model settings, but the optimal model chosen with this algorithm repeatedly generated 300 trees, well below the general rule of thumb that models should be fit with at least 1,000 trees (Elith et al., 2008). In SAHM, learning rate and tree complexity were manually parameterized using the ‘Explore’ tool to derive models with at least 1,000 trees and biologically sensible response curves. Out of the 12 original variables supplied to the BRT, 10 were used in the final historic model and only six were included in the current model.

Generalized Linear Model

The generalized linear models used in the analysis were run in the SAHM environment. Generalized linear models use a link function to relate the mean of the response variable to a linear predictor, and by doing so can handle a variety of distributions (Guisan and Zimmerman, 2000). SAHM can fit a GLM using either presence/absence or count data, which accounts for either a Poisson or binomial distribution. Generalized linear models have performed well in studies of species distribution models (Elith et al., 2006), but because they are limited in the functions they can fit, they can also generate simplified functional response curves that may not reflect realistic biological responses (Austin, 2007). The historic GLM model was fit with six final covariates and the current GLM model was fit with seven.

Principal Components Analysis (PCA)

To prepare the data for the PCA, 20,000 random background points were selected from across the study extent and variable values were extracted at each point using the “MDS Builder” tool in SAHM (Morisette et al., 2013). Additionally, data were extracted at each of the 882 occurrence localities for both the historic and current time periods. Environmental data were extracted for each time period (historic, current, and both future projections, RCP 4.5 and RCP 8.5) at the background points spanning the study area. Three separate PCA analyses were run contrasting the climatic niche shift of the pine beetle across time periods: historic to current, current to RCP 4.5, and current to RCP 8.5. The PCA analysis was conducted in R v.3.1.2 (R Core Team, 2014) using code adapted and modified from Broennimann et al. (2012). This approach applies a kernel density function to determine the smoothed density of MPB occurrences in each grid cell across the environmental space for each time period

(Broennimann et al., 2012; West et al., 2015). The density values are ordered along the PCA axes for each period’s environmental grid (the same 20,000 background points).

APPENDIX 4 – HISTORICAL MODEL OUTPUTS

Historical MaxEnt Model Results

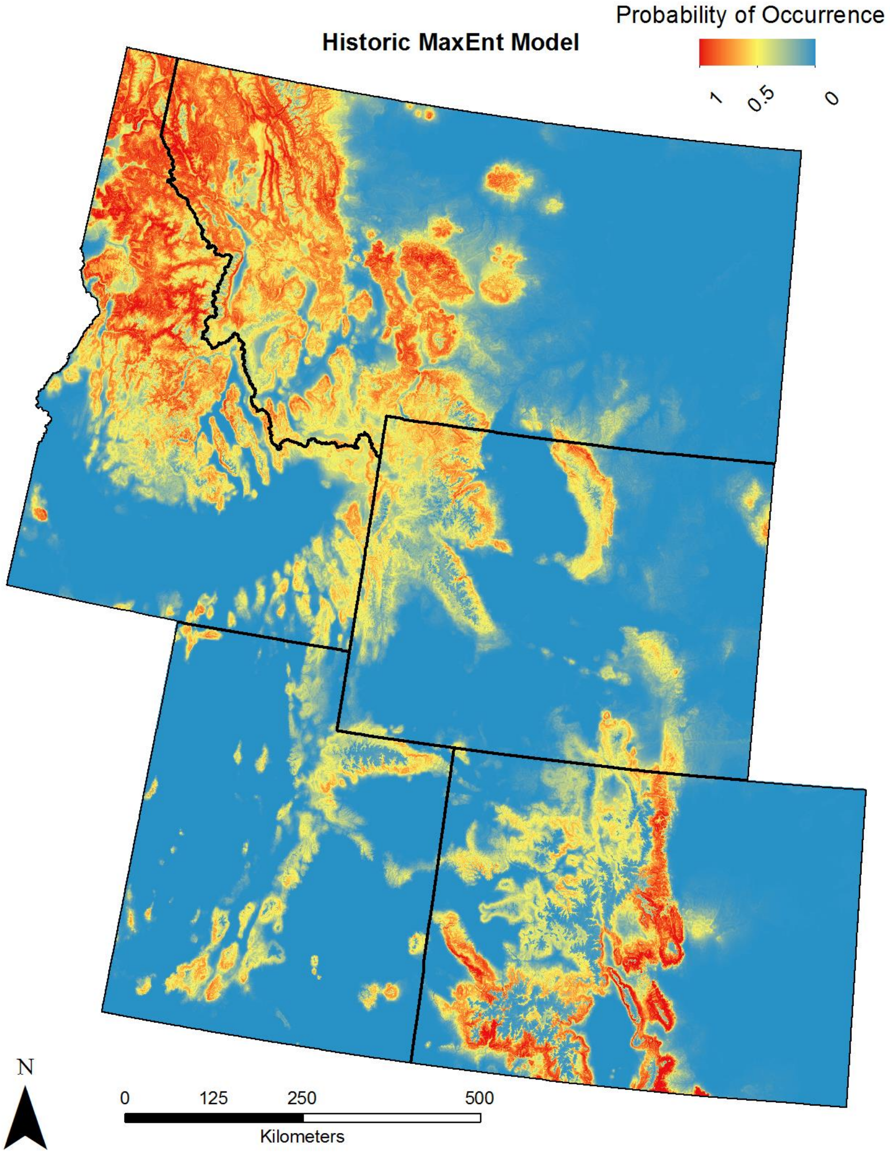

Figure 9: Historical MaxEnt predicted probability map for MPB.

Response curves

Each of the following curves represents a different model, namely, a MaxEnt model created using only the corresponding variable. These plots reflect the dependence of predicted suitability both on the selected variable and on dependencies induced by correlations between the selected variable and other variables.

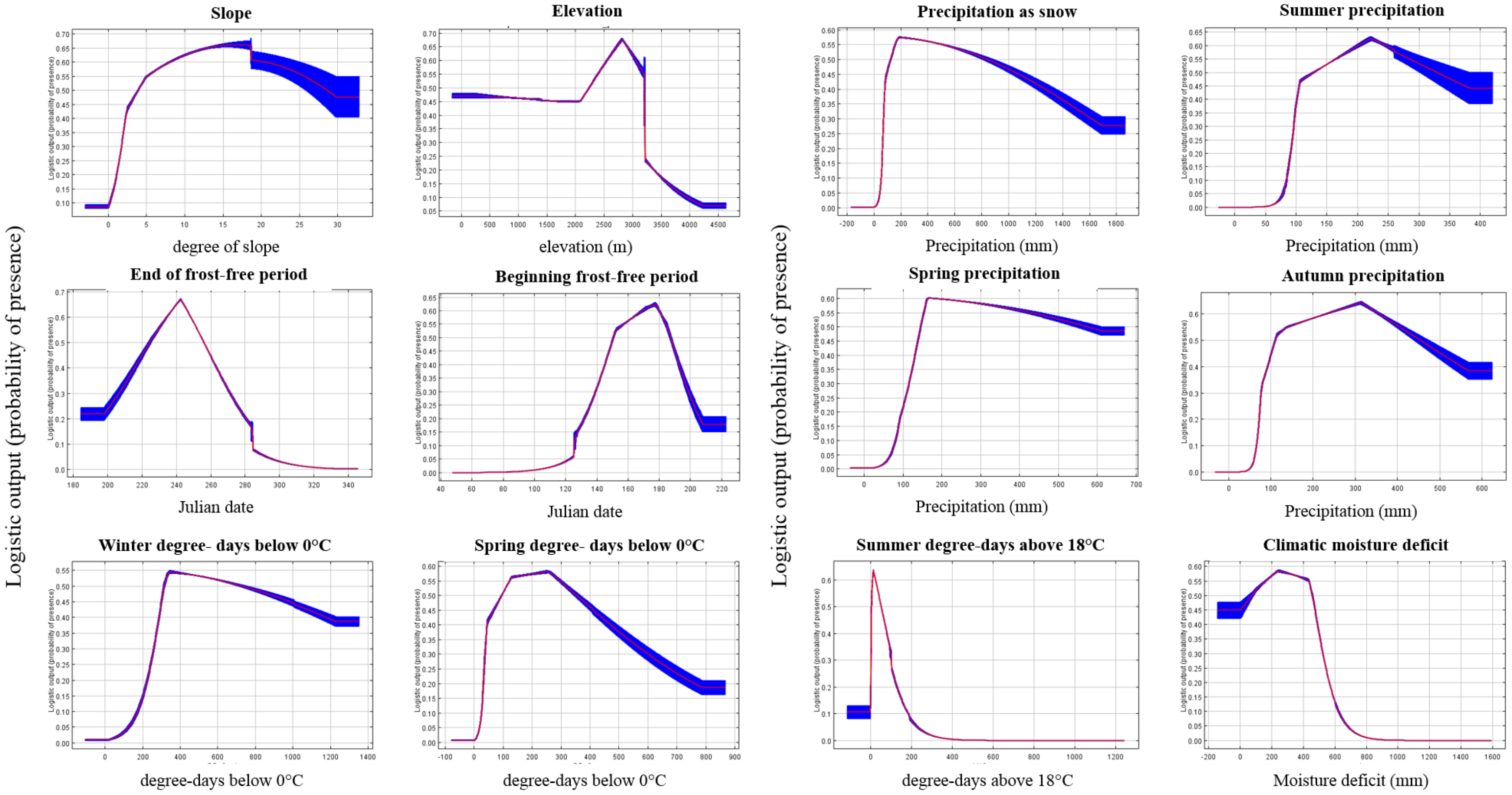

Figure 10: Historical MaxEnt species response curves.

Table 6: Historical MaxEnt variable contributions.

Variable	Percent contribution	Permutation importance
Summer precipitation	28.4	27.3
Summer degree-days above 18°C	24.1	29.8
Precipitation as snow	18.3	8.1
Climatic Moisture Deficit	7.9	2.5
Autumn precipitation	7.6	5.8
elevation	5.6	7.5
slope	3.7	3.3
Winter degree-days below 0°C	2.1	9
Spring degree-days below 0°C	0.9	2.4
end of frost-free period	0.6	1.4
beginning of frost-free period	0.6	0.2
Spring precipitation between March - May	0.2	2.6

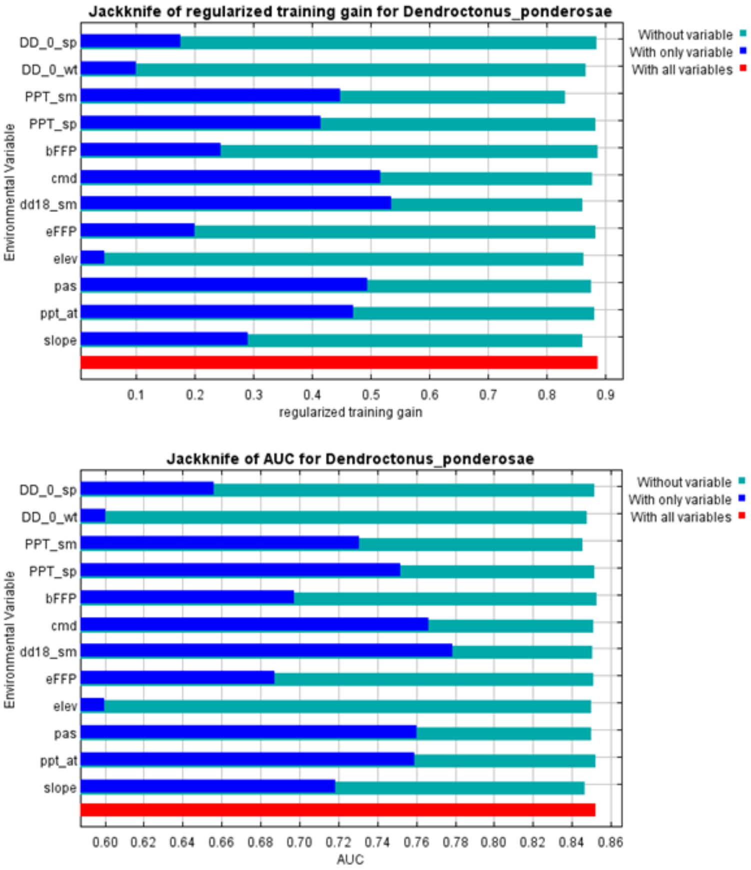

Figure 11: Relative importance of variables based on jackknife test for historical MaxEnt model. Top) training gain. Bottom) AUC.

Historical Boosted Regression Trees Model Results

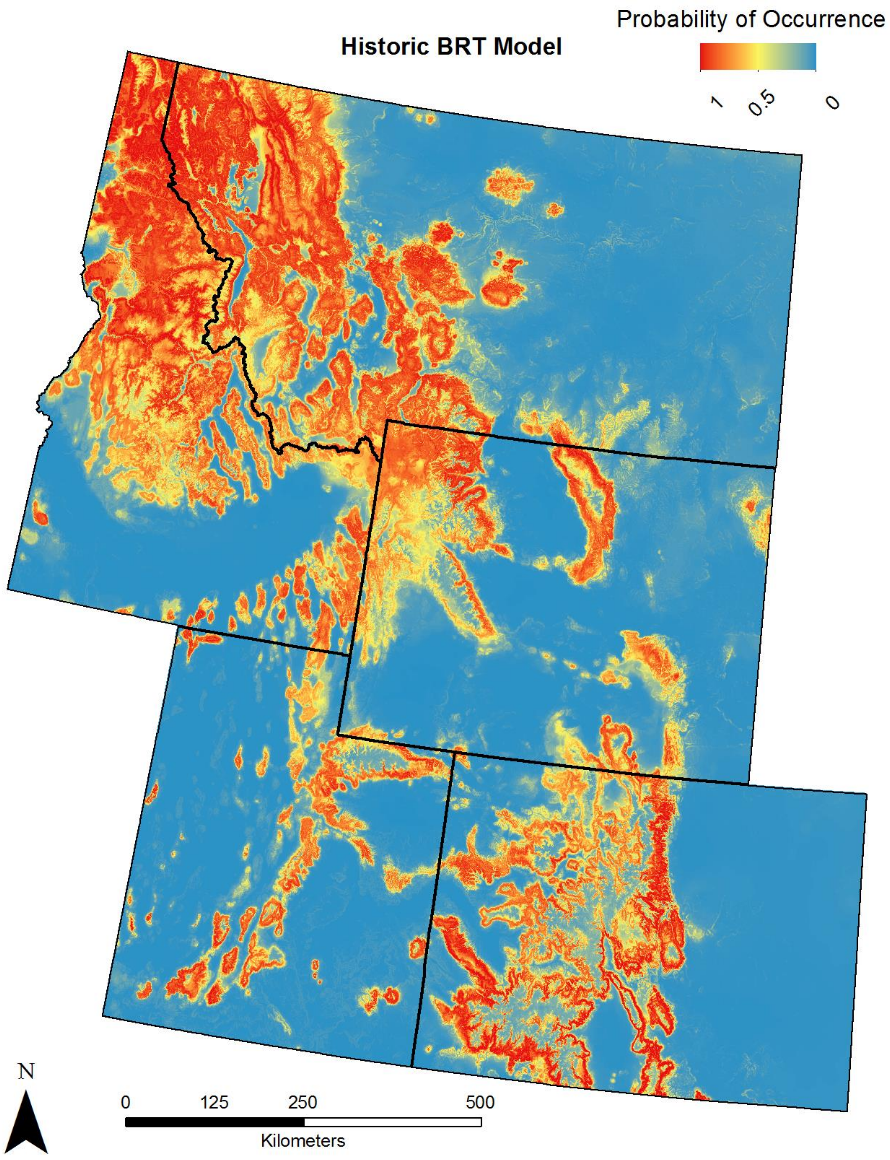

Figure 12: Historical BRT predicted probability map for MPB.

Historical BRT model results:

```
J:\Research\Modeling\SAHM\Historic\final_apply\brt_hist_BRT_5\Covari
ateCorrelationOutputMDS_.csv

      n(pres)                 =      882
      n(bkgd)                 =      20000
      n covariates considered =      12

      total time for model fitting = 14.28min


Settings:
(Averaged across available splits)

     random seed used             : 1234
     tree complexity              : 5
     learning rate                : 0.005
     n(trees)                     : 1295.45454545455
     model simplification         : cross-validation
     n folds                      : 3
     n covariates in final model  : 10
Relative influence of predictors in final model:

       Var   rel.inf
  DD18_sm 18.746163
      PAS 14.604355
      CMD 12.582452
   PPT_at 11.434956
    Slope 11.333704
   PPT_sm  9.998290
     elev  8.892213
  DD_0_wt  5.496719
   PPT_sp  3.870664
  DD_0_sp  3.040485

Important interactions in at least one split of available points:

  v1   name1 v2   name2
   6   Slope  4  PPT_at
   9    elev  4  PPT_at
   9    elev  5  PPT_sm
   9    elev  1     CMD
  10  PPT_sp  8 DD_0_wt
  10  PPT_sp  9    elev
   9    elev  3     PAS
   6   Slope  3     PAS
   8 DD_0_wt  1     CMD
  10  PPT_sp  5  PPT_sm
   9    elev  2 DD18_sm
```

```
     5  PPT_sm  1      CMD
     6   Slope  5   PPT_sm
     8 DD_0_wt  5   PPT_sm
    10  PPT_sp  6    Slope


   ============================================================

   Evaluation Statistics applied to train split:


      Correlation Coefficient      : 0.2879767
      NULL Deviance                : 1.3861 (Averaged over background
splits)
      Fit Deviance                 : 0.82768 (Averaged over background
splits)
      Explained Deviance           : 0.55839
      Percent Deviance Explained   : 40.286

     Threshold Methods based on Sens=Spec
      Threshold                    :  0.65

      Confusion Matrix:

            observed
   predicted     1     0
           1   708  4149
           0   174 15851

      AUC                          :  0.8822
      Percent Correctly Classified :  79.29796
      Sensitivity                  :  0.8027211
      Specificity                  :  0.79255
      Kappa                        :  0.1887337
      True Skill Statistic         :  0.5952711


   ============================================================

   Evaluation Statistics applied to crossValidation split:


      Correlation Coefficient      : 0.2625452  (sd 0.011833)

     Threshold Methods based on Sens=Spec
      Mean Threshold               : 0.654 (sd 0.005164)

      Confusion Matrix:

            observed
   predicted     1     0
           1   621  4139
```

```
        0   261 15861

    AUC                          :  0.84905  (sd 0.016267)
    Percent Correctly Classified :  78.92917  (sd 0.88642)
    Sensitivity                  :  0.7041113  (sd 0.057349)
    Specificity                  :  0.79305  (sd 0.0098839)
    Kappa                        :  0.1602535  (sd 0.016214)
    True Skill Statistic         :  0.4971613  (sd 0.054509)

Total time = 232.06 min
```

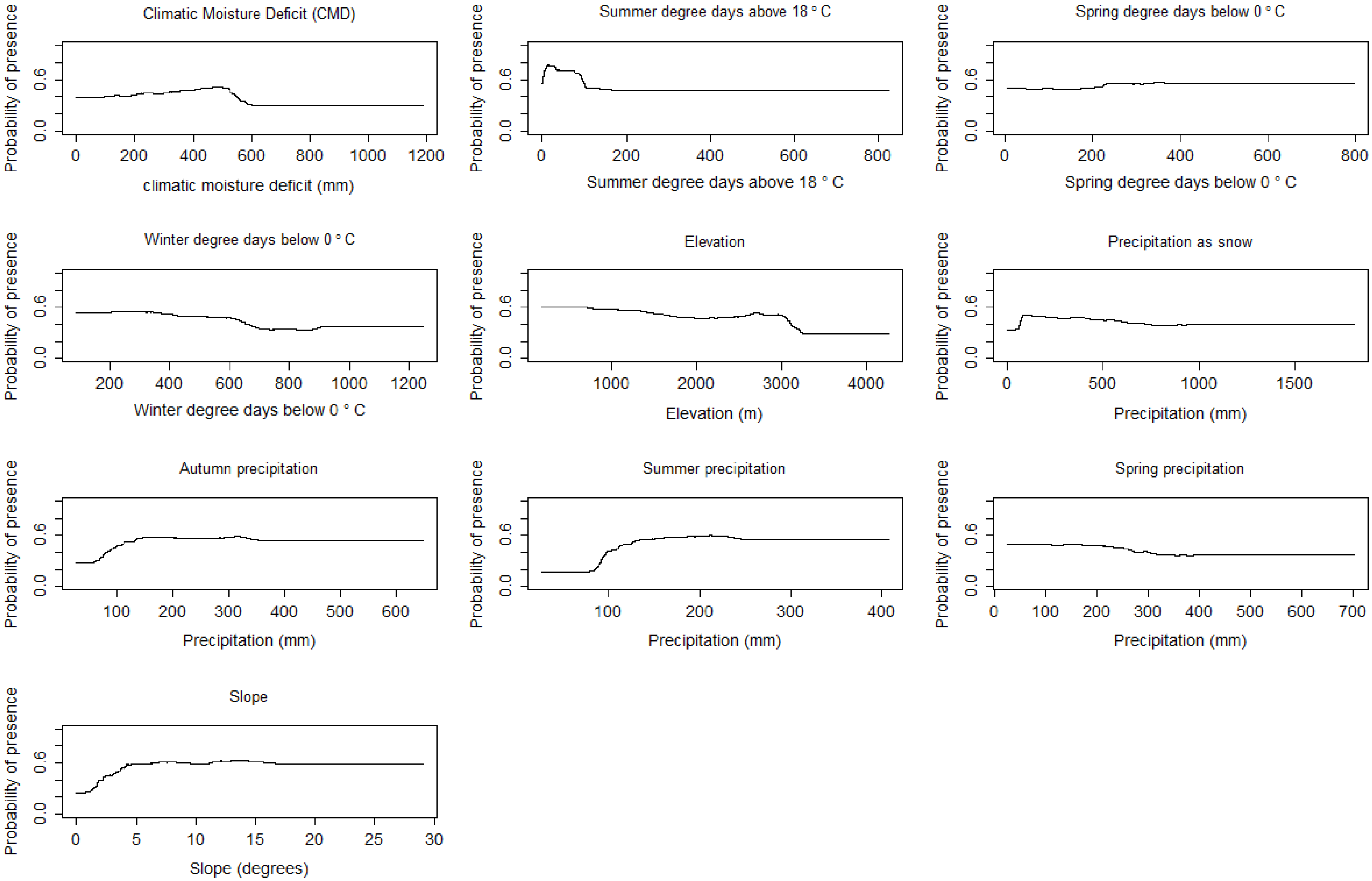

Figure 13: Historical BRT species response curves.

Historical Generalized Linear Model Results

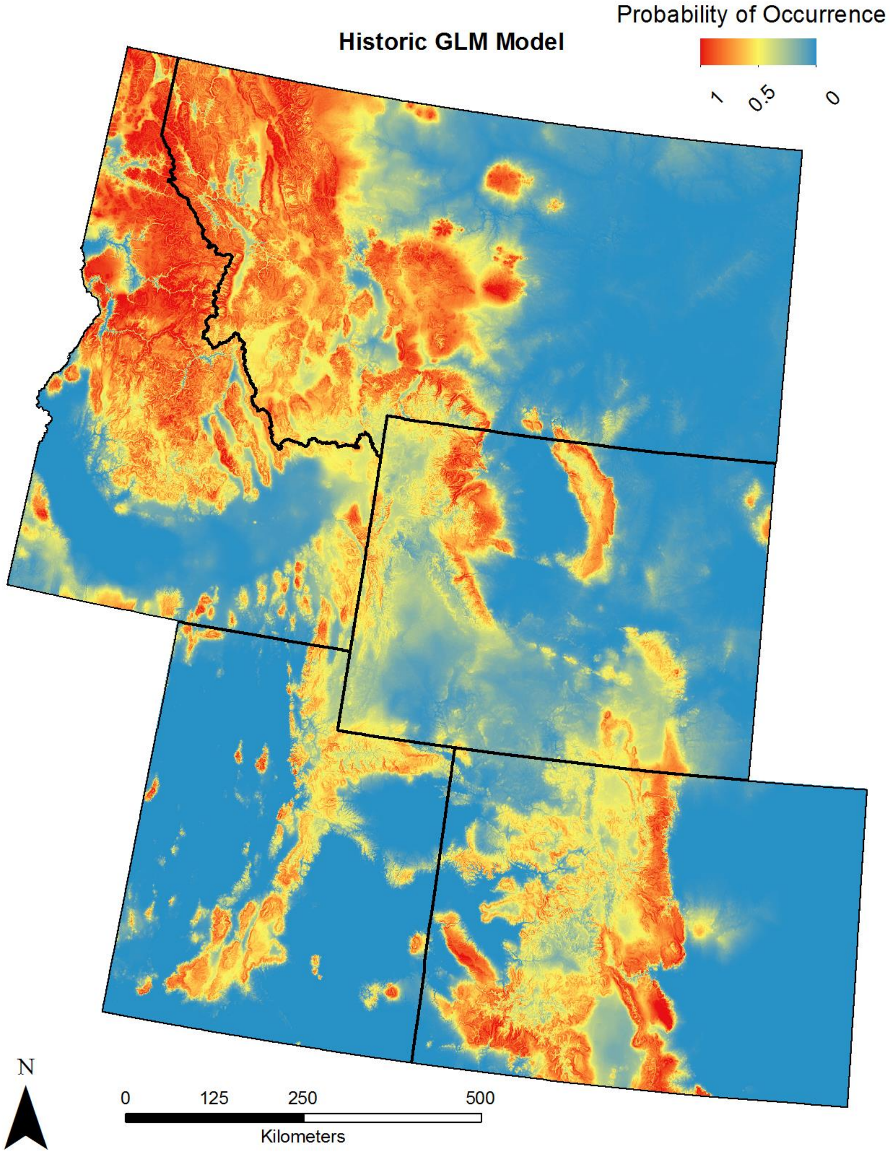

Figure 14: Historical GLM predicted probability map for MPB.

Historical GLM model results:

```
   J:\Research\Modeling\SAHM\Historic\final_apply\glm_hist_GLM_2\Covari
ateCorrelationOutputMDS_.csv

       n(pres)                  =      882
       n(bkgd)                  =      20000
       n covariates considered =      12

       total time for model fitting = 0.59min


   Settings:

       model family           : binomial
       simplification method : AIC


   Results:
       number covariates in final model   : 6

   Call:
   glm(formula = response ~ DD18_sm + DD_0_wt + PPT_sm + elev +
       Slope + PAS, family = model.family, data = dat, weights =
weight,
       na.action = "na.exclude")

   Deviance Residuals:
        Min        1Q    Median         3Q       Max
   -0.51347  -0.27985  -0.16571  -0.05659   2.87977

   Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
   (Intercept)  2.9877980  0.4182539   7.144 9.10e-13 ***
   DD18_sm     -0.0217417  0.0016642 -13.064  < 2e-16 ***
   DD_0_wt     -0.0033499  0.0004597  -7.287 3.18e-13 ***
   PPT_sm       0.0083146  0.0018570   4.477 7.55e-06 ***
   elev        -0.0005513  0.0001111  -4.963 6.94e-07 ***
   Slope        0.0792822  0.0167076   4.745 2.08e-06 ***
   PAS         -0.0009299  0.0003890  -2.390   0.0168 *
   ---
   Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

   (Dispersion parameter for binomial family taken to be 1)

       Null deviance: 2445.4  on 20881  degrees of freedom
   Residual deviance: 1777.7  on 20875  degrees of freedom
   AIC: 834.09

   Number of Fisher Scoring iterations: 6
```

```
============================================================

Evaluation Statistics applied to train split:


    Correlation Coefficient      : 0.224925
    NULL Deviance                : 0.34998
    Fit Deviance                 : 1.0792
    Explained Deviance           : -0.72918
    Percent Deviance Explained   : -208.35

  Threshold Methods based on Sens=Spec
    Threshold                    :  0.59

    Confusion Matrix:

         observed
predicted     1     0
        1   649  5194
        0   233 14806

    AUC                          :  0.8122
    Percent Correctly Classified :  74.01111
    Sensitivity                  :  0.7358277
    Specificity                  :  0.7403
    Kappa                        :  0.1290902
    True Skill Statistic         :  0.4761277


============================================================

Evaluation Statistics applied to crossValidation split:


    Correlation Coefficient      : 0.2237511  (sd 0.013817)
    NULL Deviance                : 0.34998 (sd 0.0012056)
    Fit Deviance                 : 1.0803 (sd 0.018821)
    Explained Deviance           : -0.73035 (sd 0.019337)
    Percent Deviance Explained   : -208.69 (sd 5.8793)

  Threshold Methods based on Sens=Spec
    Mean Threshold               : 0.587 (sd 0.0048305)

    Confusion Matrix:

         observed
predicted     1     0
        1   643  5206
        0   239 14794
```

```
    AUC                          :  0.81005  (sd 0.018502)
    Percent Correctly Classified :  73.92484  (sd 0.62242)
    Sensitivity                  :  0.728907  (sd 0.068058)
    Specificity                  :  0.7397  (sd 0.0061923)
    Kappa                        :  0.1268623  (sd 0.01757)
    True Skill Statistic         :  0.468607  (sd 0.06748)

Total time = 27.98 min
```

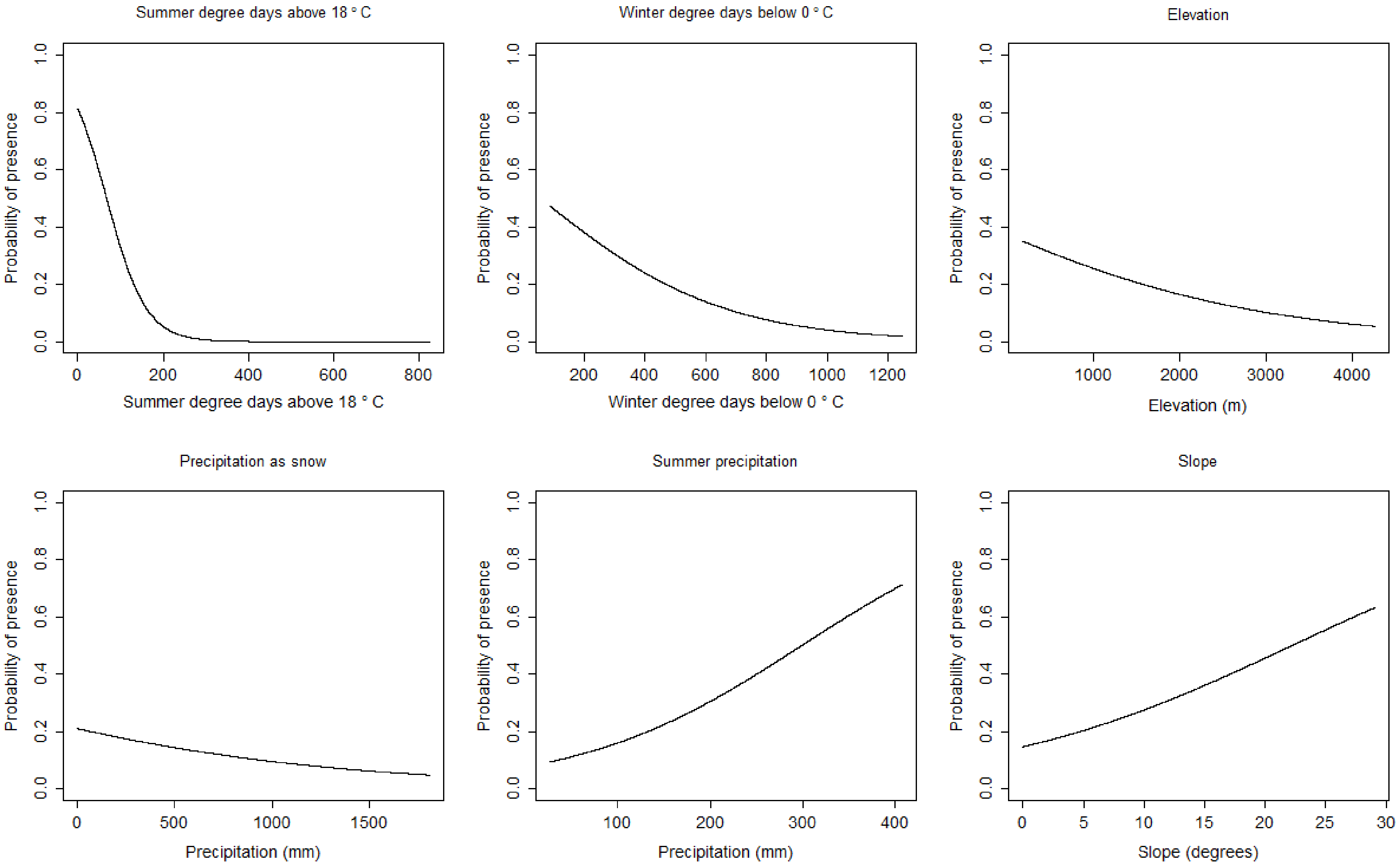

Figure 15: Historical GLM species response curves.

APPENDIX 5 – CURRENT MODEL OUTPUTS

Current MaxEnt Model Results

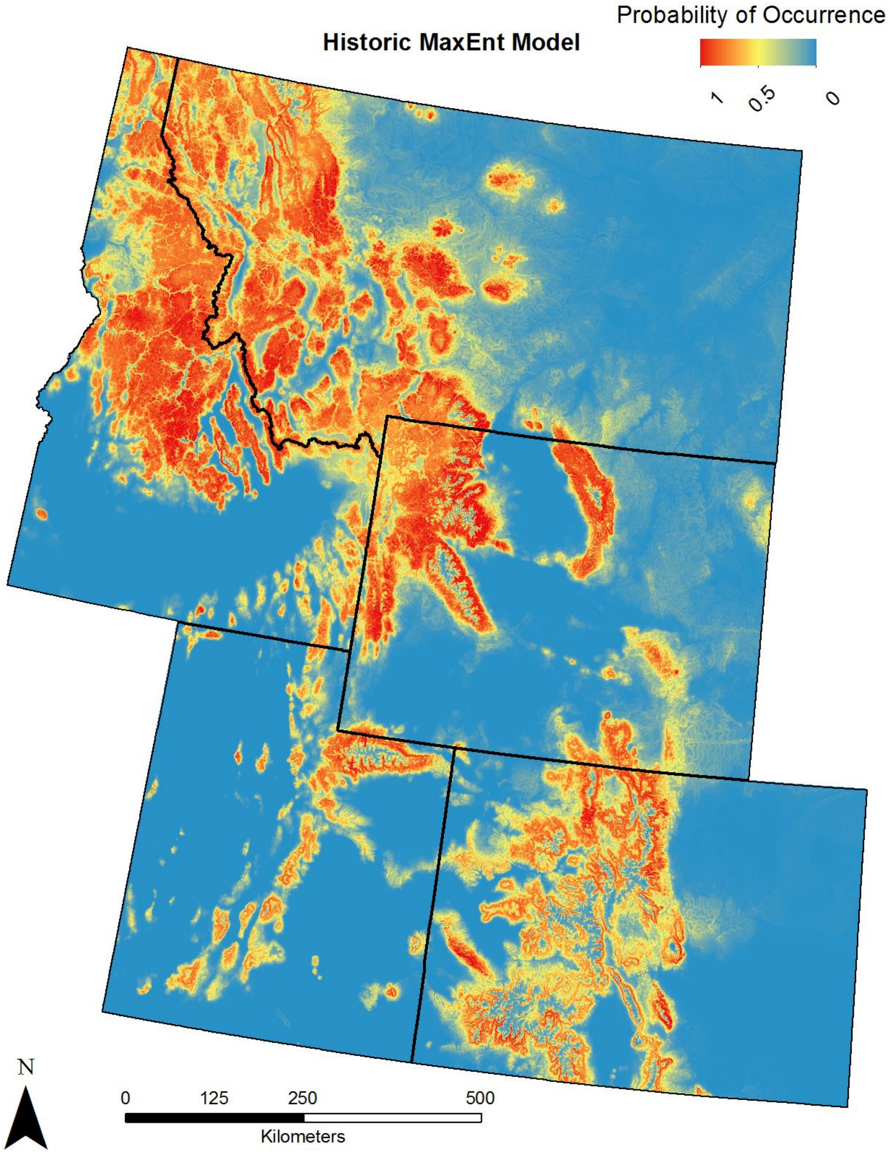

Figure 16: Current MaxEnt predicted probability map for MPB.

Response curves

Each of the following curves represents a different model, namely, a MaxEnt model created using only the corresponding variable. These plots reflect the dependence of predicted suitability both on the selected variable and on dependencies induced by correlations between the selected variable and other variables.

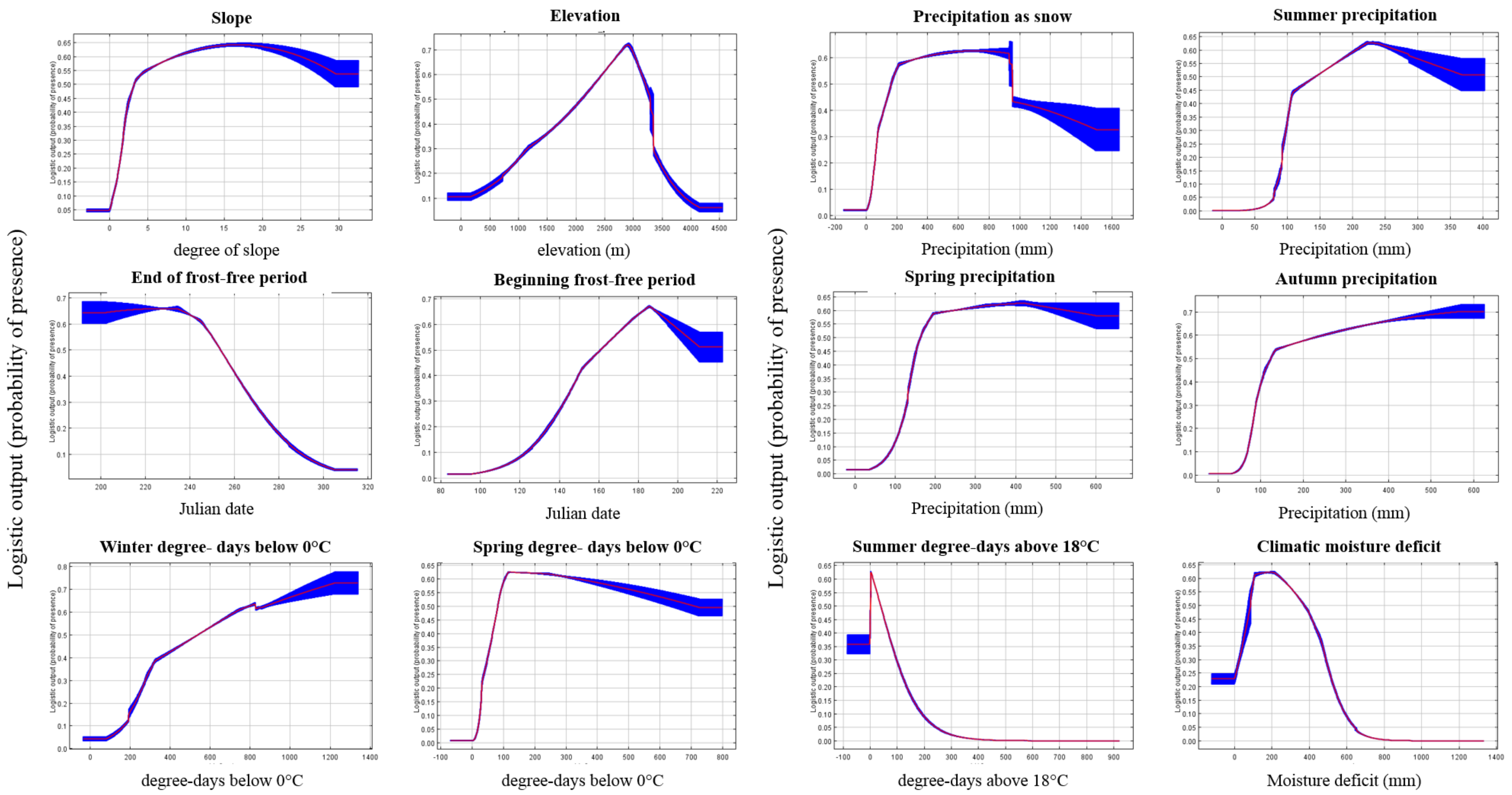

Figure 17: Current MaxEnt species response curves.

Table 7: Current MaxEnt variable contributions.

Variable	Percent contribution	Permutation importance
Climatic Moisture Deficit	41.2	22.4
Summer degree-days above 18°C	21.4	18.1
Summer precipitation	10.5	20.3
slope	10.4	11.1
Spring precipitation between March - May	5.3	0.4
elevation	5.2	6
beginning of frost-free period	1.9	0
Autumn precipitation	1.1	4.6
Precipitation as snow	0.9	6.4
Winter degree-days below 0°C	0.8	4.2
Spring degree-days below 0°C	0.7	2.6
end of frost-free period	0.5	3.8

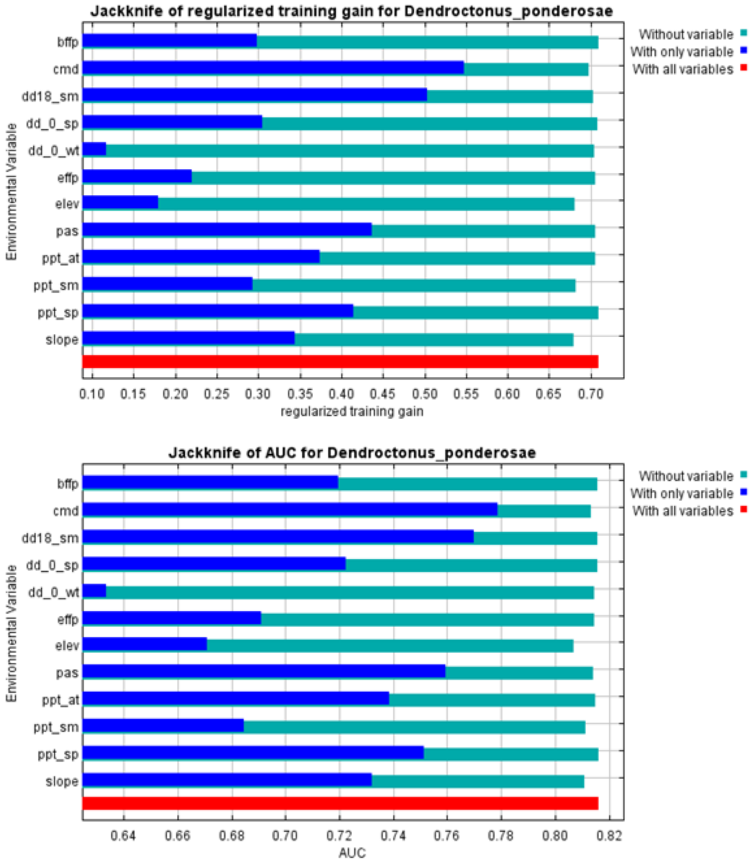

Figure 18: Relative importance of variables based on jackknife test for current MaxEnt model. Top) training gain. Bottom) AUC.

Current Boosted Regression Trees Model Results

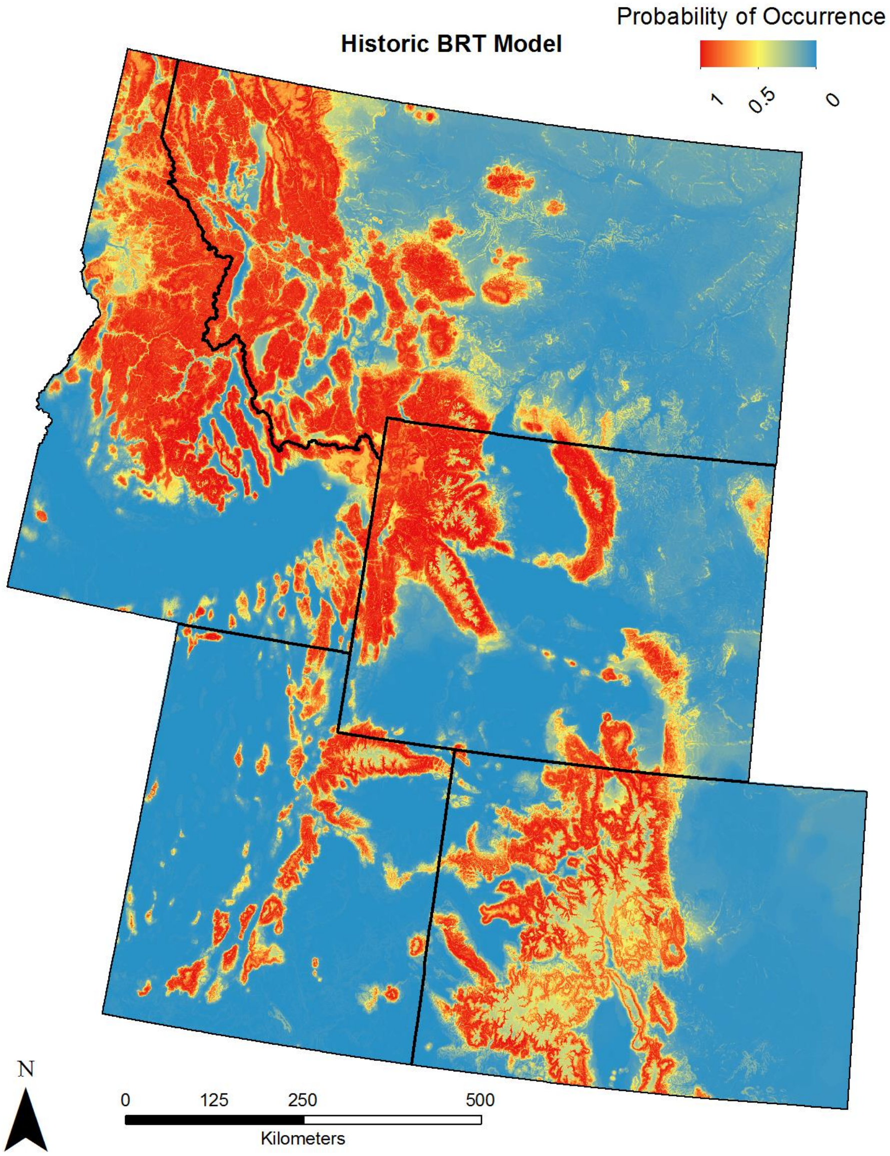

Figure 19: Current BRT probability map.

Current BRT model results:

```
   J:\Research\Modeling\SAHM\Present\final_apply2\brt_hist_BRT_1\Covari
ateCorrelationOutputMDS_.csv

      n(pres)                 =      882
      n(bkgd)                 =      20000
      n covariates considered =      12

      total time for model fitting = 10.98min


   Settings:
   (Averaged across available splits)

     random seed used             : 1234
     tree complexity              : 3
     learning rate                : 0.005
     n(trees)                     : 1009.09090909091
     model simplification         : cross-validation
     n folds                      : 3
     n covariates in final model  : 6
   Relative influence of predictors in final model:

        Var   rel.inf
        CMD 35.990234
    DD18_sm 20.883903
      Slope 13.232663
     PPT_sp 10.944816
     PPT_sm  9.527353
       elev  9.421031

   Important interactions in at least one split of available points:

    v1  name1 v2   name2
     6 PPT_sp  4   Slope
     6 PPT_sp  3    elev
     3   elev  1     CMD
     4  Slope  3    elev
     3   elev  2 DD18_sm
     5 PPT_sm  4   Slope
     5 PPT_sm  3    elev
     5 PPT_sm  1     CMD
     6 PPT_sp  5  PPT_sm


   ============================================================

   Evaluation Statistics applied to train split:
```

```
      Correlation Coefficient      : 0.2507271
      NULL Deviance                : 1.3861 (Averaged over background
splits)
      Fit Deviance                 : 0.93053 (Averaged over background
splits)
      Explained Deviance           : 0.45554
      Percent Deviance Explained   : 32.865

     Threshold Methods based on Sens=Spec
      Threshold                    :  0.65

      Confusion Matrix:

            observed
   predicted     1     0
           1   670  4705
           0   212 15295

      AUC                          :  0.8443
      Percent Correctly Classified :  76.4534
      Sensitivity                  :  0.7596372
      Specificity                  :  0.76475
      Kappa                        :  0.1526722
      True Skill Statistic         :  0.5243872


   ============================================================

   Evaluation Statistics applied to crossValidation split:


      Correlation Coefficient      : 0.2379018  (sd 0.0072098)

     Threshold Methods based on Sens=Spec
      Mean Threshold               : 0.645 (sd 0.0052705)

      Confusion Matrix:

            observed
   predicted     1     0
           1   630  4799
           0   252 15201

      AUC                          :  0.82491  (sd 0.011133)
      Percent Correctly Classified :  75.81171  (sd 0.56227)
      Sensitivity                  :  0.7143769  (sd 0.034876)
      Specificity                  :  0.76005  (sd 0.0056098)
      Kappa                        :  0.136968  (sd 0.010021)
      True Skill Statistic         :  0.4744269  (sd 0.035519)

   Total time = 153.77 min
```

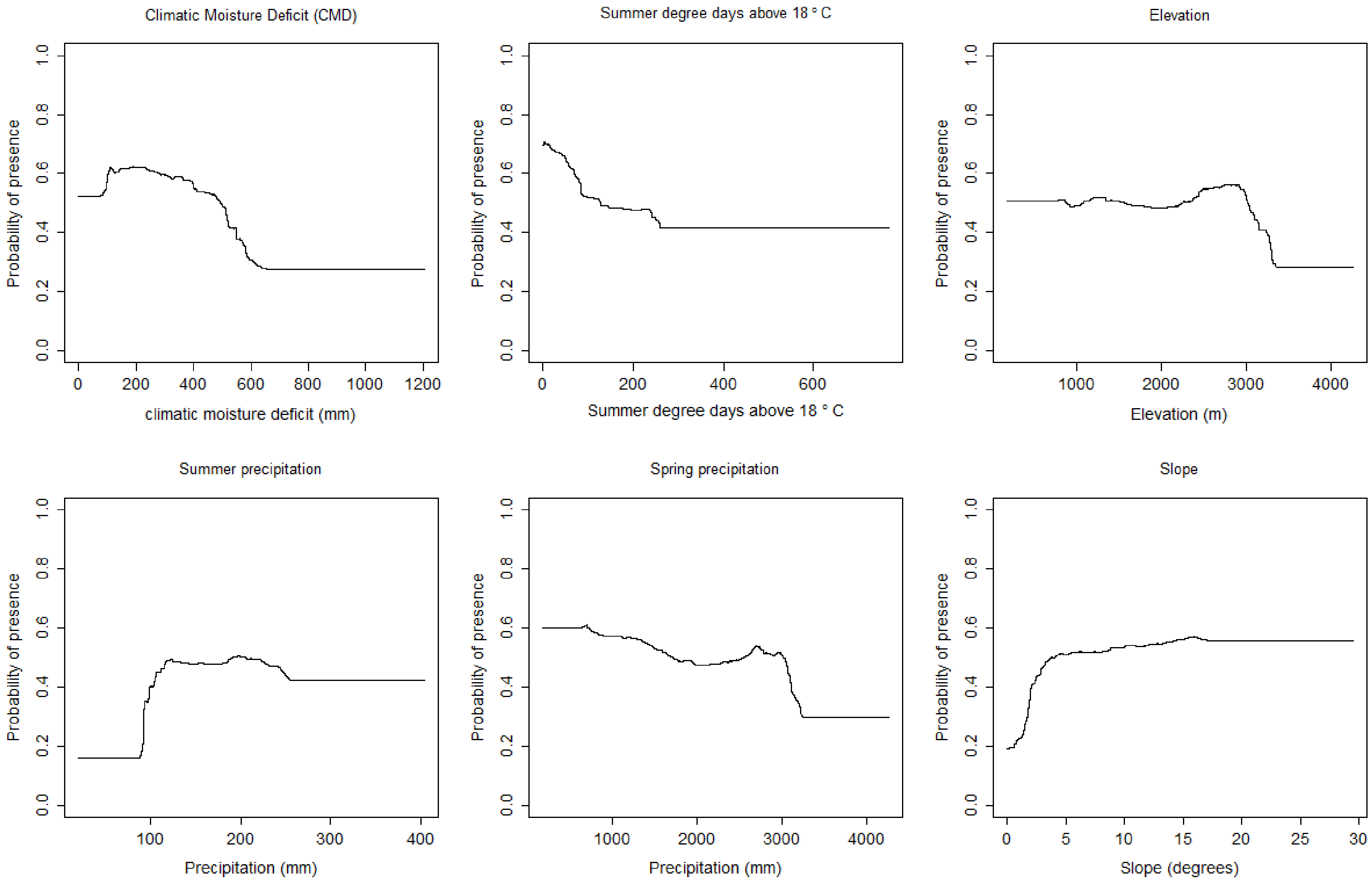

Figure 20: Current BRT species response curves.

Current Generalized Linear Model Results

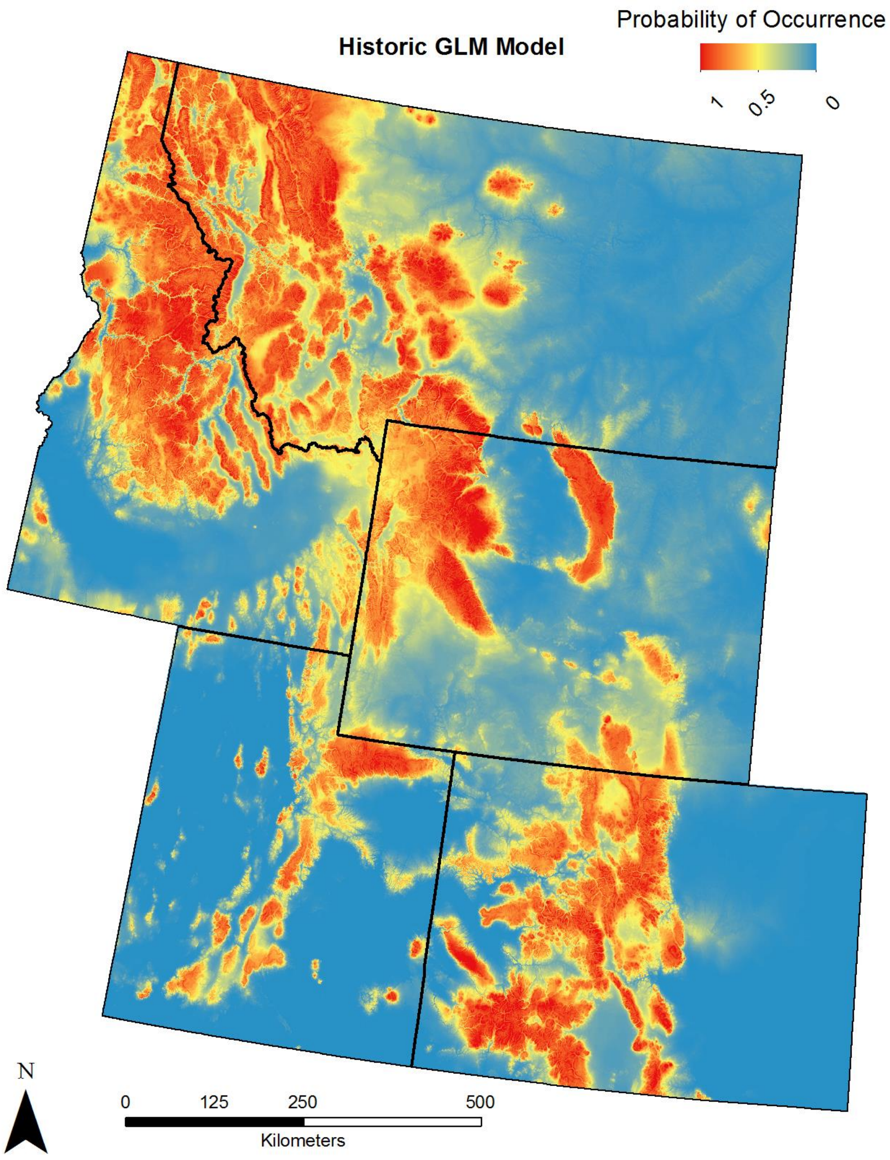

Figure 21: Current GLM predicted probability map for MPB.

Current GLM model results:

```
   J:\Research\Modeling\SAHM\Present\final_apply2\glm_hist_GLM_1\Covari
ateCorrelationOutputMDS_.csv

       n(pres)                  =      882
       n(bkgd)                  =      20000
       n covariates considered =      12

       total time for model fitting = 0.71min


   Settings:

       model family           : binomial
       simplification method : AIC


   Results:
       number covariates in final model   : 7

   Call:
   glm(formula = response ~ DD18_sm + CMD + DD_0_wt + Slope + eFFP +
       PAS + PPT_sm, family = model.family, data = dat, weights =
weight,
       na.action = "na.exclude")

   Deviance Residuals:
        Min        1Q    Median        3Q       Max
   -0.50140  -0.28693  -0.15898  -0.07468   2.45480

   Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
   (Intercept) 10.7375701  2.4068328   4.461 8.15e-06 ***
   DD18_sm     -0.0086294  0.0015361  -5.618 1.94e-08 ***
   CMD         -0.0052952  0.0009163  -5.779 7.52e-09 ***
   DD_0_wt     -0.0026091  0.0005966  -4.373 1.23e-05 ***
   Slope        0.0603926  0.0164043   3.682 0.000232 ***
   eFFP        -0.0239110  0.0075590  -3.163 0.001560 **
   PAS         -0.0009759  0.0004569  -2.136 0.032683 *
   PPT_sm      -0.0042897  0.0021570  -1.989 0.046734 *
   ---
   Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

   (Dispersion parameter for binomial family taken to be 1)

       Null deviance: 2445.4  on 20881  degrees of freedom
   Residual deviance: 1809.7  on 20874  degrees of freedom
   AIC: 853.61
```

```
Number of Fisher Scoring iterations: 6



============================================================

Evaluation Statistics applied to train split:


    Correlation Coefficient      : 0.2197055
    NULL Deviance                : 0.34998
    Fit Deviance                 : 1.0957
    Explained Deviance           : -0.74576
    Percent Deviance Explained   : -213.09

  Threshold Methods based on Sens=Spec
    Threshold                    :  0.6

    Confusion Matrix:

         observed
predicted     1     0
        1   645  5345
        0   237 14655

    AUC                          :  0.8012
    Percent Correctly Classified :  73.26884
    Sensitivity                  :  0.7312925
    Specificity                  :  0.73275
    Kappa                        :  0.1231538
    True Skill Statistic         :  0.4640425


============================================================

Evaluation Statistics applied to crossValidation split:


    Correlation Coefficient      : 0.2166679  (sd 0.010107)
    NULL Deviance                : 0.34998 (sd 0.0012056)
    Fit Deviance                 : 1.0983 (sd 0.012971)
    Explained Deviance           : -0.74828 (sd 0.013198)
    Percent Deviance Explained   : -213.81 (sd 3.997)

  Threshold Methods based on Sens=Spec
    Mean Threshold               : 0.599 (sd 0.0031623)

    Confusion Matrix:

         observed
predicted     1     0
        1   632  5316
```

```
        0   250 14684

    AUC                          :  0.79673  (sd 0.013598)
    Percent Correctly Classified :  73.34544  (sd 0.52594)
    Sensitivity                  :  0.7165603  (sd 0.042764)
    Specificity                  :  0.7342  (sd 0.0052132)
    Kappa                        :  0.120343  (sd 0.011186)
    True Skill Statistic         :  0.4507603  (sd 0.042932)

Total time = 32.21 min
```

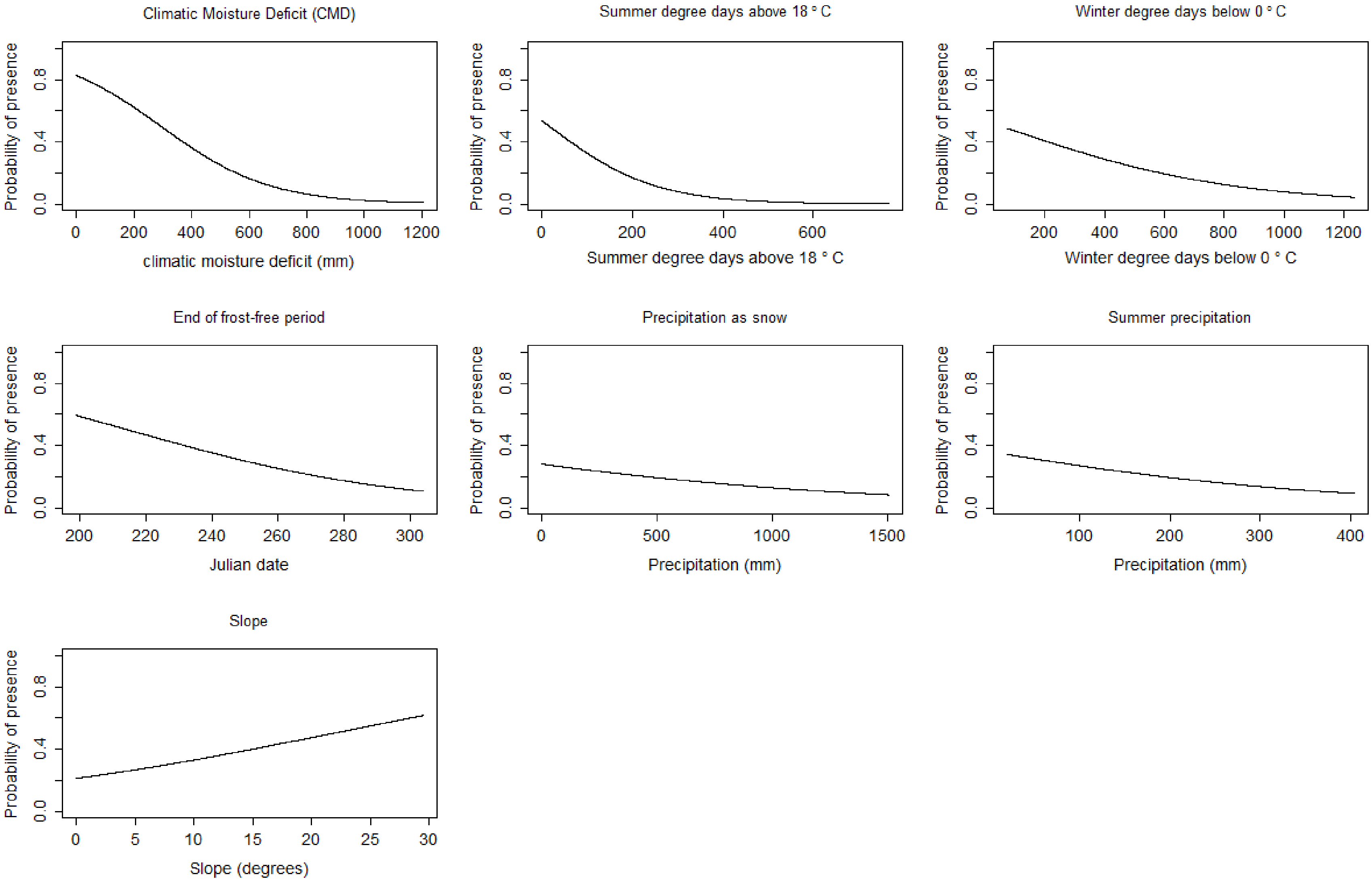

Figure 22: Current GLM species response curves.

APPENDIX 6 – LIST OF ABBREVIATIONS

ADS - aerial detection survey

AUC - area under the receiver operating characteristic curve

bFFP - beginning of the frost-free period

BRT - boosted regression trees

ClimateWNA - Climate western North America, software

CMD - climatic moisture deficit

CMIP5 - phase five of the Coupled Model Intercomparison Project multi-model data set

CO_2 - carbon dioxide

DD_0_sp - spring degree days below 0° C

DD_0_wt - winter degree days below 0° C

DD18_sm - summer degree days above 18° C

DEM - digital elevation model

eFFP - end of the frost-free period

GIS - geographic informations systems

GLM - generalized linear model

IDS - insect and disease detection survey

IPCC - Intergovernmental Panel on Climate Change

KDE - kernel density estimator

MaxEnt - maximum entropy modeling software

MESS - multivariate environmental similarity surface

MPB - mountain pine beetle

PAS - precipitation as snow (mm) between August of previous year and July of current year

PCA - principal components analysis

ppm - parts per million

PPT_at - autumn precipitation

PPT_sp - spring precipitation

PPT_sm - summer precipitation

RCP - relative concentration pathway; used to identify future climate scenarios

SAHM - Software for Assisted Habitat Modeling

SDM - species distribution model or species distribution modeling

Tmin_wt - winter mean minimum temperature (°C)

USFS - United States Forest Service

USGS – United States Geological Survey

APPENDIX 7 – DATABASE METADATA AND ORGANIZATION

Summary

The database included with this thesis contains all of the data used in the analysis including mountain pine beetle infestation data and environmental data. All data was processed in ArcGIS 10.2.1. The data are provided as an archive file format (*.zip). The data provided here would allow a user to replicate the models used in this analysis.

Projection: USA Contiguous Albers Equal Area Conic (ESRI: 102003). This is an ESRI projection equivalent to the North American Albers Equal Area Conic.

Proj4:

```
+proj=aea +lat_1=29.5 +lat_2=45.5 +lat_0=37.5 +lon_0=-96 +x_0=0
+y_0=0 +ellps=GRS80 +datum=NAD83 +units=m +no_defs
```

.PRJ File:

```
PROJCS["USA_Contiguous_Albers_Equal_Area_Conic",GEOGCS["GCS_North
_American_1983",DATUM["D_North_American_1983",SPHEROID["GRS_1980"
,6378137,298.257222101]],PRIMEM["Greenwich",0],UNIT["Degree",0.01
7453292519943295]],PROJECTION["Albers"],PARAMETER["False_Easting"
,0],PARAMETER["False_Northing",0],PARAMETER["central_meridian",-
96],PARAMETER["Standard_Parallel_1",29.5],PARAMETER["Standard_Par
allel_2",45.5],PARAMETER["latitude_of_origin",37.5],UNIT["Meter",
1]]
```

Datum: North American 1983

Raster Information: This applies to all raster data, topographic and climatic.

Extent

Top: 1453116.6009

Left: -1715671.06279

Right: -504671.062786

Bottom: -39883.3990072

Columns, Rows: 1211, 1493

Cell size (X, Y): 1000, 1000 (m)

Format: TIFF

Primary contact: Aaron Sidder, aaron.sidder@gmail.com, (720) 933-0925.

Secondary contact: Dr. Melinda Laituri, melinda.laituri@colostate.edu; 970-491-0292

Description: These data were collected from a variety of sources and are described below.

Environmental Data: The 1-km DEM was provided by Dr. Andreas Hamann of ClimateWNA and was used to download the climatic data from the ClimateWNA software at 1-km grid cell resolution. Slope and aspect were created by Aaron Sidder in ArcGIS. Historical climate data reflect climate conditions from 1951-1980, current data are from 1981-2010, and future data are from 2040-2069 from the RCP 8.5 emissions scenario. Data from the RCP 4.5 emissions scenario were not included in this database due to space constraints, but are available for download from ClimateWNA.

Historical MPB data: Historical MPB data were digitized from USFS aerial detection surveys collected from USFS regions 1, 2, and 4, and the data here date from 1960-1980. The data were acquired from regional USFS offices: Region 1- Tim Assal (U.S. Geological Survey, Graduate Degree Program in Ecology, CSU); Region 2 - Brian Howell and Justin Backsen (Forest Health Protection, U.S. Forest Service [USFS], Rocky Mountain Region); and Region 4 - Dick Halsey (Forest Health Protection, USFS, Boise Field Office). The original data was separated by year and contained information on numerous forest disturbances and pest outbreaks. This data represent the presence of the mountain pine beetle (MPB) and were extracted from the original data using the search by attributes function (DCA = 11006). After each state's MPB infestation polygons were separated, the data from all years of the study were merged into this master shapefile and all polygons were dissolved into a single feature. This polygon represents the extent of MPB infestation from 1960-1980 across the regions represented.

***Current MPB data**:* Current MPB data were collected from the USFS IDS Database: http://foresthealth.fs.usda.gov/ids. This layer combines all of the MPB aerial detection survey polygons from Colorado, Utah, Wyoming, Montana, and Idaho from the years 1997-2010. The polygons were downloaded from the USFS IDS Explorer website for the years 1997-2010, which also included data on other forest health indicators and pest outbreaks. All MPB polygons were extracted using the search by attributes function (DCA = 11006) for the years 1997-2010 to align with available climate data through Climate WNA. After each state's MPB infestation polygons were separated, the data from all five states were merged into this master shapefile and all polygons were dissolved into a single feature.

***MPB point data**:* These points were randomly generated from MPB polygons and spatially filtered so that no point is within 10 km of another point. The points were randomly generated in the Geospatial Modeling Environment (Beyer 2012) and spatially filtered in the SDM Toolbox (Brown 2014).

Study Area: The study area shapefile was clipped to the current extent and was downloaded from the U.S. Census Bureau TIGER database. The TIGER/Line shapefiles and related database files (.dbf) are an extract of selected geographic and cartographic information from the U.S. Census Bureau's Master Address File / Topologically Integrated Geographic Encoding and Referencing (MAF/TIGER) Database (MTDB).

Credits

Forest Health Protection, U.S. Forest Service; U.S. Census Bureau

ClimateWNA (Wang et al. 2012): http://climatewna.com/.

The geodatabase can be credited to Aaron Sidder, M.S. Candidate of the Graduate Degree Program in Ecology of Colorado State University and Dr. Melinda Laituri, Department of Ecosystem Science and the Natural Resource Ecology Laboratory.

Database Architecture (Note: The ESRI geodatabase contains all environmental rasters in the main folder; they are not separated into historical, current, and future datasets)

Sidder_Thesis_data

- Environmental Data
    - Historic_envtl_variables
        - Hist_Aspect.tif
        - Hist_bFFP.tif
        - Hist_CMD.tif
        - Hist_DD_0_sp.tif
        - Hist_DD_0_wt.tif
        - Hist_DD18_sm.tif
        - Hist_eFFP.tif
        - Hist_elev.tif
        - Hist_PAS.tif
        - Hist_PPT_at.tif
        - Hist_PPT_sm.tif
        - Hist_PPT_sp.tif
        - Hist_Slope.tif
        - Hist_Tmin_wt.tif
    - Current_envtl_variables
        - Curr_Aspect.tif
        - Curr_bFFP.tif
        - Curr_CMD.tif
        - Curr_DD_0_sp.tif
        - Curr_DD_0_wt.tif
        - Curr_DD18_sm.tif

        - Curr_eFFP.tif
        - Curr_elev.tif
        - Curr_PAS.tif
        - Curr_PPT_at.tif
        - Curr_PPT_sm.tif
        - Curr_PPT_sp.tif
        - Curr_Slope.tif
        - Curr_Tmin_wt.tif
    - RCP85_envtl_variables
        - Aspect_rcp85.tif
        - bFFP_rcp85.tif
        - CMD_rcp85.tif
        - DD_0_sp_rcp85.tif
        - DD_0_wt_rcp85.tif
        - DD18_sm_rcp85.tif
        - eFFP_rcp85.tif
        - elev_rcp85.tif
        - PAS_rcp85.tif
        - PPT_at_rcp85.tif
        - PPT_sm_rcp85.tif
        - PPT_sp_rcp85.tif
        - Slope_rcp85.tif
        - Tmin_wt_rcp85.tif
- Occurrence_data
    - Current_10km_albers_882.shp
    - Current_1997_2010_all_states.shp

  - Historic_10km_albers_882.shp
  - Historic_1960_1980_all_regions.shp
- Study_Area
  - study_area_albers.shp

APPENDIX 8 – VARIOUS CODE USED IN ANALYSIS

ClimateWNA Grid Processing - R

```
   ## This was written based on the tutorial from Dr. Andreas Hamann ##
   ## It worked. March 3, 2015 ##
   ## edited by Aaron Sidder ##

   ##### ----------------------------------------- #####
   ##### ----------------------------------------- #####
   # The first step of the analysis is to clip your reference DEM to
the study area and determine the cell size (resolution) of your study.
Prepare your DEM so that it matches your desired projection, grain
size, and extent.
   # Once you have created your reference DEM you will be able to use
it to download climate data from Climate WNA.
   # Save the reference DEM as an ASCII file (.asc)

   ## load R packages and set the working directory
   library(SDMTools)
   library(foreign)
   library(tools)
   setwd("J:\\Research\\DATA\\Climate_Data\\ClimateWNA_prep")

   ## Call in the ASCII reference DEM (DEM) file
   table1 = asc2dataframe("ref_elev_alb.asc")
   head(table1)
   fix(table1) #rename "var.1" to "elev"
   head(table1)
   write.csv(table1, "table1.csv", row.names=F, quote=F)

   ##### ----------------------------------------- #####
   ##### ----------------------------------------- #####
   # Read table1.csv into ArcGIS and convert to points file; it will be
projected in your initial projection.
   # Create a points shapefile and re-project as WGS84 coordinate
system (lat/long) for input into Climate WNA
   # Add the lat/long coordinates to the attribute table using the
'AddXYCoordinates' tool
   # In this example, this shapefile is named table2.shp
   # At this point, the table will have an x, y, lat, long, and elev
column

   ## Read in table2.shp to create Climate WNA input ##
   table2 <- read.dbf("table2.dbf")
   head(table2)
   table3 <- table2[,c(1,2,5,4,3)] # reorganize column to preferred
order for ClimateWNA
```

```
   head(table3)
   fix(table3) #col headers as y, x, lat, long, elev
   head(table3)
   write.csv(table3, "table3.csv", row.names=F, quote=F)

   # table 3 will be used in Climate WNA in its current format.  You
can rename it to be more specific if needed. ##

   ## read the Climate WNA output CSV and export to individual ASCII
files ##
   setwd("J:\\Research\\DATA\\Climate_Data\\ClimateWNA_output\\table3_g
lobalmean_rcp85_2050_seasonal")
   list.files(pattern="*.csv")
   table4 =
read.csv("table3_GlobalMean_rcp85_r1i1p1_2050s_seasonal.csv",
nrows=1808030, header=T)
   head(table4)
   ncol(table4)
   ## annual
   #table4 = table4[,c("y","x","bFFP", "eFFP", "PAS", "CMD")]

   ## seasonal
   table4 = table4[,c("y","x","Tmin_wt", "DD_0_wt", "DD_0_sp",
"DD18_sm", "PPT_at", "PPT_sm", "PPT_sp")]

   ## all
   #table4 = table4[,c(1,2,6:ncol(table4))] # can break into batches of
50 for large datasets
                                            # (6:50, 51:100,
101:ncol(table4))

   head(table4)
   # update output directory with folder name
   dataframe2asc(table4, outdir=paste(getwd(), sep=""))
   rm(list = ls()) # clear environment
   gc() # garbage collect, drops memory allocation
```

SAHM – PARC data preparation - R

```
   library(tools)

   ## Future climate variables ##

   rcp45 = list_files_with_exts(dir = 
"J:/Research/Modeling/ENV_data/Future_variables_rcp45/TIF/PARC_RCP45_f
uture_ref_elev_alb", exts = "tif")

   rcp85 = list_files_with_exts(dir = 
"J:/Research/Modeling/ENV_data/Future_variables_rcp85/TIF/PARC_RCP85_f
uture_ref_elev_alb", exts = "tif")

   ## Write future predictor lists to CSV
   head(rcp45)
   rcp45.predictor <- cbind(rcp45, "0", "Bilinear", "Mean")
   colnames(rcp45.predictor) <- c("PARCOutputFile", "Categorical", 
"Resampling", "Aggregation")
   head(rcp45.predictor)
   nrow(rcp45.predictor) # should be 14
   write.csv(rcp45.predictor, 
"J:/Research/Modeling/SAHM/rcp45_predictors.csv", row.names=F, 
quote=F)

   head(rcp85)
   rcp85.predictor <- cbind(rcp85, "0", "Bilinear", "Mean")
   colnames(rcp85.predictor) <- c("PARCOutputFile", "Categorical", 
"Resampling", "Aggregation")
   head(rcp85.predictor)
   nrow(rcp85.predictor) # should be 14
   write.csv(rcp85.predictor, 
"J:/Research/Modeling/SAHM/rcp85_predictors.csv", row.names=F, 
quote=F)

   write.csv(rcp45.predictor, 
"J:/Research/Modeling/SAHM/rcp45_predictors.csv", row.names=F, 
quote=F)

   head(rcp85)
   rcp85.predictor <- cbind(rcp85, "0", "Bilinear", "Mean")
   colnames(rcp85.predictor) <- c("PARCOutputFile", "Categorical", 
"Resampling", "Aggregation")
   head(rcp85.predictor)
   nrow(rcp85.predictor) # should be 14
   write.csv(rcp85.predictor, 
"J:/Research/Modeling/SAHM/rcp85_predictors.csv", row.names=F, 
quote=F)
```

ENMeval for Maxent regularization – R

```
   ## ------ Run ENMevaluation on Maxent ------ ##
   ## ------ February 3, 2015 ------ ##
   ## 882 Historical data points @ 10 km min distance ##

   rm(list=ls())
   setwd("J:\\Research\\Modeling\\ENV_data\\Present_variables")
   list.files(pattern =".asc")

   ## load libraries ##
   library(ENMeval)
   library(raster)

   ## format data points ##
   pts <-
read.csv("J:\\Research\\Modeling\\FieldData\\Pres_10km_maxent.csv",
header = TRUE)
   head(pts)
   occ <- pts[,c(2,3)]
   head(occ)
   nrow(occ) # nrow = 882

   ## call in rasters and create rasterStack ##
   # create raster layers from variables
   aspect <-
raster("J:\\Research\\Modeling\\ENV_data\\Present_variables\\aspect.as
c")
   bFFP <-
raster("J:\\Research\\Modeling\\ENV_data\\Present_variables\\bFFP.asc"
)
   cmd <-
raster("J:\\Research\\Modeling\\ENV_data\\Present_variables\\cmd.asc")
   DD_0_sp <-
raster("J:\\Research\\Modeling\\ENV_data\\Present_variables\\DD_0_sp.a
sc")
   DD_0_wt <-
raster("J:\\Research\\Modeling\\ENV_data\\Present_variables\\DD_0_wt.a
sc")
   dd18_sm <-
raster("J:\\Research\\Modeling\\ENV_data\\Present_variables\\dd18_sm.a
sc")
   eFFP <-
raster("J:\\Research\\Modeling\\ENV_data\\Present_variables\\eFFP.asc"
)
   elev <-
raster("J:\\Research\\Modeling\\ENV_data\\Present_variables\\elev.asc"
)
   PAS <-
raster("J:\\Research\\Modeling\\ENV_data\\Present_variables\\PAS.asc")
```

```
   ppt_at <-
raster("J:\\Research\\Modeling\\ENV_data\\Present_variables\\ppt_at.as
c")
   PPT_sm <-
raster("J:\\Research\\Modeling\\ENV_data\\Present_variables\\PPT_sm.as
c")
   PPT_sp <-
raster("J:\\Research\\Modeling\\ENV_data\\Present_variables\\PPT_sp.as
c")
   slope <-
raster("J:\\Research\\Modeling\\ENV_data\\Present_variables\\slope.asc
")
   tmin_wt <-
raster("J:\\Research\\Modeling\\ENV_data\\Present_variables\\tmin_wt.a
sc")

   # create rasterStack from rasters
   env <- stack(aspect, bFFP, cmd, DD_0_sp, DD_0_wt, dd18_sm, eFFP,
elev, PAS, ppt_at, PPT_sm, PPT_sp, slope, tmin_wt, native=TRUE)
   nlayers(env)

   ## create Maxent arguments
   RM <- c(seq(1.0, 4.0, 0.5))
   fc <- c("LQH", "LQHP", "LQHT","LQHPT")
   #fc <- c("LQH", "LQHP", "LQHT", "LQHPT")

   ## run ENMevaluate ##
   enmeval_results <- ENMevaluate(occ, env, RMvalues=RM, fc=fc,
method="randomkfold", kfolds=5, n.bg=20000, overlap=FALSE,
bin.output=TRUE, clamp=TRUE)

   #enmeval_results <- ENMevaluate(occ, env, RMvalues=RM, fc=fc,
method="randomkfold", kfolds=5, n.bg=10000, overlap=TRUE,
categoricals=c("nlcd2006forrcls"),bin.output=TRUE, clamp=TRUE)

   ## view results
   enmeval_results

   ### See table of evaluation metrics
   results <- enmeval_results@results
   write.csv(results,
"J:\\Research\\Modeling\\Maxent\\Maxent_present\\ENMeval\\ENMeval_resu
lts3.csv", row.names=F, quote=F)
   enmeval_results@overlap


   ## plot results
   par(mfrow=c(2,2))
   eval.plot(enmeval_results@results, legend.position="topright")
   eval.plot(enmeval_results@results, "Mean.AUC",
legend.position="right")
```

```
   eval.plot(enmeval_results@results, "Mean.AUC.DIFF",
variance="Var.AUC.DIFF")
   eval.plot(enmeval_results@results, "Mean.ORmin",
legend.position="right")
```

MPB Polygon, Select by Attribute, all regional codes – Python

```
   #-------------------------------------------------------------------
------------
   # Name:        Select by Attributes batch
   # Purpose:     To separate out mountain pine beetle polygons by
agent using extract by attributes for all coding methods in National
forest units.
   #
   # Author:      Aaron Sidder
   #
   # Created:     16/10/2014
   # Copyright:   (c) asidder 2014
   # Licence:     <your licence>

   # ***** BEFORE YOU RUN THIS CODE, CHANGE mywspace, outputFolder,
new, lyr, selectLayer *****
   # ***** Change forest and folder each run *****
   # ***** UPDATE README.TXT IN THE FOLDERS AS YOU GO *****
   #-------------------------------------------------------------------
------------

   #Import system module
   import arcpy
   from arcpy import env
   from arcpy.sa import *

   # Define workspace
   mywspace =
"K:\\Research\\DATA\\ADS_data\\Historic\\R2_ADS_Historical_Maps\\Digit
ized Data\\SanJuan"
   arcpy.env.workspace = mywspace

   #Set overwrite on/off
   arcpy.env.overwriteOutput = "TRUE"

   #
**********************************************************************
*********
   # Coding Method 1
   #
**********************************************************************
*********

   # Output folder - Coding method 1
   outputFolder1 = "K:\\Research\\DATA\\Historic\\SanJuan\\MPB_Code1"

   try:
   # This can be ListDatasets, ListFeatureClasses, ListFiles,
ListRasters, ListTables, or ListWorkspaces
```

```
       featureclassList1 = arcpy.ListFeatureClasses("*","POLYGON") # or
use (,"ALL")
       for featureClass1 in featureclassList1:
               print featureClass1
               new1 = outputFolder1 + "\\" + featureClass1[0:-4] +
"_MPB" + ".shp"  # new = mywspace + "\\" + featureClass + "_MPB"
#this is to be used for a gdb
               print new1
               lyr1 = arcpy.MakeFeatureLayer_management(featureClass1,
"lyr1" )
               print lyr1
               select_lyr1 =
arcpy.SelectLayerByAttribute_management(lyr1, "NEW_SELECTION", "
\"CODE\" LIKE '5%' OR \"CODE\" LIKE '6%' OR \"CODE\" LIKE '7%' ")    #
SelectLayerByAttribute_management (in_layer_or_view, {selection_type},
{where_clause})
               final1 = arcpy.CopyFeatures_management(select_lyr1,
new1)   # CopyFeatures_management (in_features, out_feature_class,
{config_keyword}, {spatial_grid_1}, {spatial_grid_2},
{spatial_grid_3})
               print final1
               arcpy.SelectLayerByAttribute_management(lyr1,
"CLEAR_SELECTION")

   # Folder will contain all shapefiles but will have empty files
deleted manually.
   # Note: when reviewing the files note that 5/6/7 that indicate MPB
also are used to code different beetles in coding method 2.  If a
color is associated with the code
   # then the coding is method 2 (e.g. 5-blue, black hills beetle light
intensity)

   except:
       print "Failure for coding method 1."
       print arcpy.GetMessages()

   else:
       print "coding method 1 completed successfully."

   #
*****************************************************************
*********
   # Coding Method 2
   #
*****************************************************************
*********

   # Output folder - Coding method 2
   outputFolder2 = "K:\\Research\\DATA\\Historic\\SanJuan\\MPB_Code2"

   try:
```

```
   # This can be ListDatasets, ListFeatureClasses, ListFiles,
ListRasters, ListTables, or ListWorkspaces
       featureclassList2 = arcpy.ListFeatureClasses("*","POLYGON") # or
use (,"ALL")
       for featureClass2 in featureclassList2:
               print featureClass2
               new2 = outputFolder2 + "\\" + featureClass2[0:-4] +
"_MPB" + ".shp"  # new = mywspace + "\\" + featureClass + "_MPB"
#this is to be used for a gdb
               print new2
               lyr2 = arcpy.MakeFeatureLayer_management(featureClass2,
"lyr2" )
               print lyr2
               select_lyr2 =
arcpy.SelectLayerByAttribute_management(lyr2, "NEW_SELECTION", "
\"CODE\" LIKE '4%' ")    # SelectLayerByAttribute_management
(in_layer_or_view, {selection_type}, {where_clause})
               final2 = arcpy.CopyFeatures_management(select_lyr2,
new2)   # CopyFeatures_management (in_features, out_feature_class,
{config_keyword}, {spatial_grid_1}, {spatial_grid_2},
{spatial_grid_3})
               print final2
               arcpy.SelectLayerByAttribute_management(lyr2,
"CLEAR_SELECTION")

   # Folder will contain all shapefiles but will have empty files
deleted manually.

   except:
       print "Failure for coding method 2."
       print arcpy.GetMessages()

   else:
       print "coding method 2 completed successfully."


   #
**********************************************************************
*********
   # Coding Method 3-19: Coding beginning with "MPB..." instead of
numbers.
   #
**********************************************************************
*********

   # Output folder - Coding method 3
   outputFolder3 =
"K:\\Research\\DATA\\Historic\\SanJuan\\MPB_Lettering_Code"

   try:
   # This can be ListDatasets, ListFeatureClasses, ListFiles,
ListRasters, ListTables, or ListWorkspaces
```

```
       featureclassList3 = arcpy.ListFeatureClasses("*","POLYGON") # or
use (,"ALL")
       for featureClass3 in featureclassList3:
               print featureClass3
               new3 = outputFolder3 + "\\" + featureClass3[0:-4] +
"_MPB" + ".shp"  # new = mywspace + "\\" + featureClass + "_MPB"
#this is to be used for a gdb
               print new3
               lyr3 = arcpy.MakeFeatureLayer_management(featureClass3,
"lyr3" )
               print lyr3
               select_lyr3 =
arcpy.SelectLayerByAttribute_management(lyr3, "NEW_SELECTION", "
\"CODE\" LIKE 'MPB%' ")    # SelectLayerByAttribute_management
(in_layer_or_view, {selection_type}, {where_clause})
               final3 = arcpy.CopyFeatures_management(select_lyr3,
new3)   # CopyFeatures_management (in_features, out_feature_class,
{config_keyword}, {spatial_grid_1}, {spatial_grid_2},
{spatial_grid_3})
               print final3
               arcpy.SelectLayerByAttribute_management(lyr3,
"CLEAR_SELECTION")

   # Folder will contain all shapefiles but will have empty files
deleted manually.

   except:
       print "Failure for coding method 3."
       print arcpy.GetMessages()

   else:
       print "     Coding method 3 completed successfully."
       print "     All processes completed successfully."

   #
**********************************************************************
*********
   # End of Script
   #
**********************************************************************
*********
```