Locating Stable Sites: Climate Refugia in the Southern Appalachians

Alexander Harper Nelson
Clemson University, alecnelson24@gmail.com

Follow this and additional works at: https://tigerprints.clemson.edu/all_theses

Recommended Citation
https://tigerprints.clemson.edu/all_theses/2723

This Thesis is brought to you for free and open access by the Theses at TigerPrints. It has been accepted for inclusion in All Theses by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.
LOCATING STABLE SITES: CLIMATE REFUGIA IN THE SOUTHERN APPALACHIANS

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Wildlife and Fisheries Biology

by
Alexander Harper Nelson
August 2017

Accepted by:
Dr. Robert F. Baldwin, Committee Chair
Dr. Kyle Barrett
Dr. Patrick Gerard
Dr. Joan L. Walker
ABSTRACT

Climate change continues to be one of the most challenging threats to global biodiversity and species persistence. In response, conservation design researchers and applied practitioners have recently begun to call for the identification of critical areas of stable climatic and environmental conditions that may preserve the platform of current climate dynamics, and promote the adaptation and dispersal of diverse taxa across the landscape. Due to their historically buffered and resilient features, climate refugia are considered valuable conservation targets that may function as robust bastions for climatically-sensitive endemic species.

In this thesis research, I have worked to define the potential stability of refugia areas within the topographically-complex, and biologically-diverse Southern Appalachian Mountain region. Specifically, I developed a methodology that used regional-scale geographic and climate data in a geospatial context. To develop this novel application of multivariate control chart-based techniques to assess the stability of climate patterns at each site, I extracted temperature, precipitation, and topographic data from sites in the region, upon which statistical models of stable refugia on the mountain landscape were constructed. The resulting output was incorporated into a mapped representation of possible sites for conservation implementation. While many important academic research refinements are possible over the next several years, this research framework and these results will be of immediate value in prioritizing critical areas for rare and threatened species in this region. These technological advances will help inform geospatial modeling work in landscape-scale conservation design.
I would like to thank my thesis committee for their outstanding guidance, including Rob Baldwin, Kyle Barrett, Patrick Gerard, and Joan Walker. I would also like to recognize Paul Leonard at the Arctic Landscape Conservation Cooperative, Travis Belote at The Wilderness Society, Peter White at the University of North Carolina-Chapel Hill, and John Montroy for their critical discourse in defining the considerations and approaches conducted through this project. I thank my peers in Wildlife and Fisheries Biology and around Clemson University for their inspirational and educational dialogue; the faculty and administration of the Clemson Department of Forestry and Environmental Conservation; and the inspiration provided by the work done by the Appalachian Landscape Conservation Cooperative.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TITLE PAGE</td>
<td>i</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vi</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Historically Employed Methods</td>
<td>8</td>
</tr>
<tr>
<td>Conservation Applications</td>
<td>16</td>
</tr>
<tr>
<td>II. RESEARCH DESIGN AND METHODS</td>
<td>19</td>
</tr>
<tr>
<td>Purpose of the Study</td>
<td>19</td>
</tr>
<tr>
<td>Explanation of Research Design</td>
<td>19</td>
</tr>
<tr>
<td>Summary of Methods</td>
<td>22</td>
</tr>
<tr>
<td>Results</td>
<td>28</td>
</tr>
<tr>
<td>Discussion</td>
<td>30</td>
</tr>
<tr>
<td>Understanding the Model Results</td>
<td>30</td>
</tr>
<tr>
<td>Justification for and Overview of Methodological Decisions</td>
<td>31</td>
</tr>
<tr>
<td>Connections and Conclusions</td>
<td>32</td>
</tr>
<tr>
<td>Further Implications and Recommendations for Further Research</td>
<td>33</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>38</td>
</tr>
<tr>
<td>A: Final R Statistical Software Script</td>
<td>39</td>
</tr>
<tr>
<td>B: Generalized Linear Mixed Model Statistical Output</td>
<td>53</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>54</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Climate Variables</td>
<td>23</td>
</tr>
<tr>
<td>2.2</td>
<td>Topographic Variables</td>
<td>24</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Flowchart of Analysis</td>
<td>20</td>
</tr>
<tr>
<td>2.2</td>
<td>Study Area Map</td>
<td>22</td>
</tr>
<tr>
<td>2.3</td>
<td>Example of Control Chart</td>
<td>27</td>
</tr>
<tr>
<td>2.4</td>
<td>Map of Percent Likelihood of Refugia</td>
<td>29</td>
</tr>
</tbody>
</table>
CHAPTER ONE
INTRODUCTION

The survival of many rare and endemic species residing in mountainous areas will be threatened by the changing nature of Earth’s climate and habitats in the coming century. One goal for conservation design is to develop the methods and priorities that will allow these species and landscapes to persist through the numerous pressures already evident. Some species have the capacity to disperse to more favorable areas, while others may be able to adapt to new conditions, but the likelihood of these responses being adequate can be uncertain or speculative. While species-specific approaches to conservation design have been successful in the past, the challenges we face may require a more landscape-based approach. One such approach is to identify the characteristics of areas that promote long-term biodiversity and long-term efficacy of the biosphere. “Climate refugia” have been proposed as a model to effectively prioritize land for reserves and linkages, as well as maintain populations of specialized endemics (Reside et al., 2013; Harrison and Noss, 2017). The National Climate Change Adaptation Research Facility defines refugia as, “…areas within the landscape which are naturally buffered from extreme variation in environmental conditions…” (NCCARF, 2015). Hence they are, as Keppel et al. (2012) states, “…habitats that components of biodiversity retreat to, persist in and can potentially expand from under environmental conditions.” Refugia have historically supported the persistence of species biodiversity and high biotic density over millennia and changing climates, and are hypothesized to have served as bastions for rare endemic species due to their environmentally stable nature (Keppel et al., 2012; Harrison and Noss, 2017).
Therefore, it is believed that certain types of modern-day refugia have the potential to harbor native biodiversity and support landscape-scale distribution of multiple taxa. Through better identification and understanding of climate refugia, these areas can serve as foundations for conservation practices on a variety of terrestrial landscapes. Several approaches have been utilized to produce models of refugial areas, but none have emerged as a clear standard for defining what constitutes “climate change refugia”. Here, I will demonstrate a method of assessing site stability over time and mapping that to a suite of topographic predictors across the landscape. This coarse filter approach provides a foundation for current prioritizations within the Southern Appalachian region, and future research of refugia identification.

Climate refugia offer a vital landscape-scale function over a significant time period due to their unique buffering capacity. Evidence suggests that a wide range of organisms have utilized these areas, which have offered shelter for temperature-limited species during periods of glaciation such as the Last Global Maximum (LGM) (Keppel et al., 2012). Refugia such as buffered valleys or mountaintop areas have the capacity to serve as sanctuaries for species that are in the process of adapting to or retreating from unfavorable climates and many such sites still contribute this ecosystem function. In this way, current species and ecosystem assemblages are maintained, supporting more robust landscape diversity in areas such as sheltered valleys or poleward-facing slopes (Morelli et al., 2016). Modeling approaches to identify these climate refugia require measurements of environmental characteristics (e.g., topography, land use) and climatic projections (e.g., local measurements and downscaled Global Climate Models) (Kearney and Porter, 2009;
Ashcroft, 2010; Reside et al., 2013). Modern computer software and data analytics, the latest in statistical tools, and better satellite and surface-based mapping of geographies and weather, are allowing rapid advancements in this field.

Climate refugia are thought to feature characteristics that make them ideal conservation targets for the preservation of species that prefer less extreme, more stable sites under increasing environmental disturbance. These areas may feature climatic conditions up to 6°C cooler than the surrounding landscape, thereby reducing the variety of extreme heat stresses on these organisms (NCCARF, 2015). Temperature-related climatic stability is likely a continuation of several-million-year trends in areas, which have high species endemism and biodiversity over a significant paleological time-scale (Reside et al., 2013; Harrison and Noss, 2017). From these stable refugia areas, climatically-displaced species may better adapt to new environmental conditions and potentially even expand. However, the implications for the long-term viability of endemic populations by protecting and managing these sites and buffering from future change are not fully understood. The degree to which refugial areas can promote dispersal and habitation is not known, nor is it clear which set of species or size of populations may find them most beneficial and in what capacity. Additionally, by constraining the focus of management to a few key areas and features of the landscape, surrounding communities of neighboring organisms may shift to encroach or avoid them. One ideal outcome of this research project is that, through the identification of these characteristically stable systems, climate refugia may be better understood and incorporated into conservation management activities based on stakeholder priorities.
Climate refugia are delimited by the characteristics that define their physical structure and, subsequently, their biological function. Geological formations and elevation gradients have long been hypothesized to be the primary source of long-term expression of biodiversity and distribution of species in Appalachian refugia (Whittaker, 1956; Cogbill and White, 1991). Metrics of topography and soil types have previously been proposed as a method of identifying “land facets” or “geophysical settings” to support future biodiversity (Anderson and Ferree, 2010; Beier and Brost, 2010; Anderson et al., 2014). The Nature Conservancy conducted an assessment of “resilient sites”, which are defined by considering landscape diversity and site connectedness (Anderson et al., 2014). Resilient sites are distinguished based on the geophysical setting and permeability of the area to the movement of organisms. Anderson et al. developed a methodology to demonstrate that high landform heterogeneity will support diverse ecosystem processes, as well as tested whether this metric performed better than measurements of climate patterns in explaining biodiversity on the landscape. Species may demonstrate different attributes of resiliency to climate, but at the fundamental landscape level, overall patterns of biodiversity are indicated by their “resilient sites” model (Anderson et al., 2014). Climate refugia models, on the other hand, are often defined by the expression of climatic interactions based on topographic factors. While each model framework is investigating the role of topography in establishing priority areas on the landscape, “resilient sites” focus strictly on geophysical factors, while climate refugia integrates the relationship of measured and projected climate into the final analysis. In both cases, spatial heterogeneity of topoclimatic systems has been shown to have significant potential to buffer organisms
against changing climatic pressures, particularly for enabling organisms to shift their distributions in response to climatic oscillation in montane systems (Ackerly et al., 2010; Harrison and Noss, 2017).

Refugia are important for species that are less able to disperse freely or those that require specific bioclimatic conditions, such as on mountaintops or in coves in the southern Appalachians. Both of these ecosystems harbor exceptional species diversity for endemic cold-adapted species, though are likely to respond differently to climate change. While mountaintop refugia may become threatened by shifting climatic conditions and therefore unsuitable for current communities (losing community stability at a higher rate), coves are expected to maintain a greater degree of persistence of extant species over the same time period (Dobrowski, 2011). Appalachian cove sites such as Fortune’s Cove, VA and Sequatchie Valley, TN feature sheltered and concave slope surroundings, which accumulate nutrients and moisture to support abundant plant and animal communities, as well as high local structural complexity. Climate refugia may serve as current and future targets for core habitat preservation, as well as areas to be connected through corridors, contributing to local ecosystem structure in a fragmented landscape (NCCARF, 2015). Efficient identification and protection of these climate refugia is, therefore, more critical now than ever, as their number, size, and degree of habitat connectivity is certainly under pressure from anthropogenic forces.

The Southern Appalachians of the eastern Unites States are a prime example of a region with both a paleoecological history conducive to forming climate refugia and high species endemism and richness, especially in areas such as the Cumberland Plateau and
Blue Ridge Mountains (Loehle, 2007; Jenkins et al., 2015). The variety of soils, microclimates, and topographic elements contribute to niches for a variety of climate-sensitive local species (Loehle, 2007). The climatic niches in these mountain systems appear to have functioned as refugia for species for the past 18,000 years, as the last glacial advance began to recede at the beginning of the Holocene, leading to significant changes in climate and vegetation across the region (Morin and Unger 1997). The prevalence of diverse surviving paleoendemics through the glacial period, as well as neoendemic species generated through the opening of unique climatic zones, provide historical evidence of the fostering effect of the topographic and climatic complexity of this area (Loehle, 2007; White, 2008; Harrison and Noss, 2017). The mesic sites, such as coves, northern slopes, and stream valleys have historically served as refuge sites for sensitive species as their niches changed and adapted to the overall drying effects of the glacial retreat (Cogbill and White, 1991; Lohle, 2007). Recognizing the role of the historic refugial sites presents a solid starting point for approaching the complex task of identifying them in the present and potentially projecting their future.

Both the spatial scale and specific defining criteria must be considered for evaluating a climate refugium impact on species and in conservation planning. Many “macro-refugia”, such as mountains, valley systems, or forest networks, feature large heterogeneous areas of current stable conditions over a landscape-scale geographic range, while “micro-refugia”, such as valley floors, hollows, or hillsides, are more compact local protected spaces (Olsen et. al, 2012; NCCARF, 2015). Therefore, classification of refugia may be on the scale of several kilometers or a few meters,
making their definition and identification challenging. For the purposes of this study, the “refugia” to be identified are at the macro-scale and therefore summarize more detailed climate and habitat information. A distinction of the macro-refugia is that larger, more heterogeneous areas have historically and in many cases will continue to buffer against changing conditions, and be more suitable to larger-bodied and more widely-distributed species. Future research should in parallel define smaller refugia, which may be more precarious and transient, but importantly can also be home to smaller, rarer, and micro-climate specialized populations of animals and plants (Ashcroft, 2010).

Refugia identification should also be contingent on the priority species or taxa, based on their functional niches and dispersal capabilities. While the “arena” of ecosystem functions may support organisms based on geophysical heterogeneity and soil diversity (Beier and Brost, 2010), different species will continue to vary in their dispersal and distributional response to changing climatic conditions, shifting the extent and grain over which their conservation should be considered (Lawler et al., 2009). Many critical species may be spatially constrained in climate refugia due to their specific habitat requirements, and therefore more susceptible to reduction in their dependent areas or food sources (Ashcroft, 2010). To assess the conservation benefits of refugia, managers may consider prioritizing areas by both their vulnerability, as well as irreplaceability relative to the priority species. Refugia must be considered from both a climatic stability, and habitat suitability perspective, with the methodology chosen to define priority areas based on a species or ecosystem focus, as well as the relevant time-scale of the conservation design (Ashcroft, 2010). Multiple climate adaptation strategies may be required to capture the
variation between species and the rates at which they may be able to disperse or adapt to changing environmental conditions, which could alter the process for determining the scale of each refugial analysis. Modeling techniques grounded in both climatic and habitat-based criteria will be able to more accurately and efficiently estimate the location and attributes of refugia resistant to climate change by incorporating the wide range of potentially contributing factors at multiple scales.

**Historically Employed Methods**

An Overview of Methodological Considerations

In the emerging field of identifying and classifying climate refugia, several methods have been proposed to characterize and model these environmental areas across a range of spatial and temporal scales. Methodologies for identifying climate refugia fall into two primary categories, pattern-based or process-based (Keppel et al., 2012).

The *pattern-based methods* of identifying refugia stem from organism-specific biogeographic patterns, which utilize paleobiology, ecology, and genetics of species of interest to set parameters, for where refugia may exist in the present. Using data such as pollen records, macrofossils, phylogeography, ecological traits of species such as limited dispersal and longer life spans, and spatially-derived genetic evidence allows one to detect the extent of refugia from historical origins (Gavin et al., 2014). Ecological niche models and other similar correlative methods have also been used to infer the location of likely refugia by combining predicted distributions of suitable conditions for target species in a historical context (Waltari et al., 2007). Biogeographical and fossil records have defined
refugial areas through the Last Glacial Maximum (LGM) for southern thermophilic plant species in the Appalachian region (Gonzales et al., 2008). Such pattern-based models have significant promise to inform the assessment of potential valuable habitat on a biologically-specific scale, but are limited by an inability to define precise causal relationships, account for all influential factors affecting a system, or to be extrapolated beyond the limits of the model (Araújo and Peterson, 2012). Pattern-based models have had some success in defining priority conservation areas, most notably from the Australian National Climate Change Adaptation Research Facility described in the comprehensive report by Reside, et al. in 2013. The extensive assessment of the entire Australian continent focused primarily on models of species endemism and distribution, identifying areas that featured greater value at present and into multiple future climate scenarios. Additionally, they noted the influence of Pleistocene stability on the diversity of current taxa, as well as how seasonal drought and monsoon effects define the importance of protected sites on an annual basis. While such regional analyses gave a broad picture to aid systematic conservation prioritization, they noted that future work must include downscaling observations of species-patterns to the local level for the greatest confidence in the model results (Reside et al., 2013).

Utilizing Process-based Methodologies

*Process-based methods* involve identifying the broad-scale processes that have a high probability of supporting refugia habitat, based on multi-dimensional factors of physical geography and environmental processes. Data such as regional topography and climatic data are selectively merged to more accurately describe: levels of radiation and
shading, remotely-sensed digital elevation models, other variations such as geodiversity and soil heterogeneity, comparisons to downscaled climate change models, distribution maps of limiting resources, LiDAR data measuring canopy height and vegetative structure, and history and evidence of disturbance (e.g. glaciations, high-intensity fires, major weather events), leading to succession may all be used in modeling refugia sites (Ashcroft et. al, 2012; Anderson et al., 2014). Analyses of climate velocity, the rate of spatial shift of climate patterns over time (Loarie et al., 2009), and climatic decoupling, anomalous variation in temperatures or precipitation relative to the surrounding atmospheric conditions (Lesser and Fridley, 2015), has been considered relevant for understanding montane-climate relationships. Lower climate change velocity has recently been connected to areas of high species endemism, with rugged topography as an important contributing factor to the “slower” nature of climatic patterns on the landscape (Harrison and Noss, 2017). However, climate velocity has been recently rebutted as an ideal measure for understanding climate change exposure in mountainous regions, since it underestimates exposure where climate trajectories cross many dissimilar areas (Dobrowski and Parks, 2016). This metric is nevertheless promising as one method of quantifying the pressures faced by climate-sensitive organisms in future climate scenarios (Loarie et al., 2009; Harrison and Noss, 2017).

Refugia are often described as resulting from high spatial heterogeneity and diversity, which generate conditions for forming a variety of climate niches. These spaces are typically associated with topographically complex features, such as mountain ranges and deep valleys, as well as more subdued heterogeneity in the landscape such as derived
from geological classes, landform, and latitude (Ashcroft et. al, 2012; Morelli et al, 2016). Landscape-scale topographic diversity appears to drive biological response to local climatic variation and has been demonstrated to correlate strongly with the distribution and diversity of numerous animal and plant species, as well as estimated landscape resilience (Anderson and Ferree, 2010; Beier and Brost, 2010; Anderson et al., 2014). Refugia can also be temporally defined, with different species responding in their adaptation or dispersal over time; e.g., temperature, water availability, highlighting that refugia are both dynamic in time and seasonality (Keppel et al., 2012). Analyses of where these climatically-stable areas may form have been based on both historical inference (based on existing refugia from the Last Glacial Maximum), and observation of current conditions; as well as projecting such areas into future climatic scenarios (Waltari et al., 2007; Walker et al., 2009; Roberts and Hamann, 2016).

Process-based models are often constructed by defining the structure of the areas on a topographic fabric and relating the climatic characteristics of a site to associated environmental variables, (Curtis et al., 2014; Lesser and Fridley, 2015). Identifying climate refugia from a spatial-topographic perspective therefore requires that the position and landscape structure be considered for its role in shaping the resulting ecosystems. Typical surface metrics include: elevation, slope, topographic position, aspect, solar insolation, profile and planiform curvature, and topographic ruggedness index, which all characterize the form of the landscape, as well as some underlying soil features (Beier and Brost, 2010; Anderson et al., 2014). Geophysical heterogeneity also drives many properties of patterns of diversity, providing evidence that conserving the stage based on
geologic factors has the potential to support future biodiversity (Anderson and Ferree, 2010). One notable recent example of this type of analysis was the identification of refugial meadows connected throughout the varied topography of the Sierra Nevada, California (Maher et al., 2017). Using models of historical climatological data and comparisons to future climate projections, the estimated differences in site conditions were determined; sites features with less change over time were classified as refugial. The relationship between these sites and key topographic factors such as elevation and regional connectivity were tested, resulting in the conclusion that refugia tended to exist in this system at higher elevations and greater connectivity, as well as the indication that such refugial sites would become more scarce into the future (Maher et al., 2017).

While there are numerous critical components in understanding refugial structure, in this research project I have primarily focused on how relatively climatically stable a site has been over time, and how that characteristic may be an indicator of the presence of high priority habitat for temperature-sensitive species.

Modeling Stability of Refugial Sites

The stability of a site’s climate relative to the surrounding environment is considered to be one of the most critical factors in understanding whether refugia will persist over time, thereby supporting a range of environmentally-sensitive organisms. For the purposes of this study, stable sites were considered to be those that experience few extreme environmental fluctuations when compared to other sites at the same time of year. Climatic stability has historically been assessed through multiple methods, though
primarily calculated by comparing overall change between pairs of past and present conditions (Iwamura et al., 2010; Ashcroft et al., 2012; Belote et al., 2012).

There are several challenges to quantifying how stable a site has been or is going to be, both in how the variance of a site over time can be quantified, and determining the relevant climatic variables. Sites may have relatively similar climatic characteristics over time, which may make determining a threshold of “unusual” variation challenging. To this point, stability can also be measured by investigating the inverse condition, namely in determining the relative instability or variation of temperature and precipitation in the system (Epstein and McCarthy, 2004). This allows one to infer which sites that would be unlikely to serve as refugia, due to their higher inconsistency of climate over time. By calculating the inverse of stability, we can predict which sites have the potential to continue to maintain a steady climatic condition and the topographic factors influencing those systems.

Several methods have been proposed to calculate a metric of climate stability, including comparing warming rates or changes in variance, and measuring climate velocity (Epstein and McCarthy, 2004; Loarie et al., 2009). However, no single calculation has been widely accepted that can sufficiently address all factors involved in measuring stability. For this project, a straightforward multivariate measure was chosen to estimate the variation in climatic processes, and subsequently the stability of sites across the landscape. One unexplored measurement is through the use of quality control charts.
Quality control charts have been used since the 1920s in the manufacturing and engineering fields to track the instability (or variation) of a system with repeated measures, as well as statistically identify the states of the system and demonstrate processes that are performing outside of established limits. Each system is consolidated into a series of values, which are used to form upper and lower control limits based on the standard deviation of the data, after which all values that exceed these limits are denoted as such. This statistical measure offers a clear visualization of the variability of observations, as well as a comparable statistic for understanding differences in a process over time. If a system has measurements within these control limits, it can be considered more predictable and stable than one that more frequently exceeds those limits (Radziwill, 2015). The quality control chart methodology has been used in the ecology field to analyze environmental variables; however, such a modeling approach is new to the question of mapping defined climate refugia (Nugraha et al., 2017). The benefits of using the control chart approach in this study, includes a measurement system that allows for repeatable measure on multiple climate variables, the ability of control charts to account for multivariate measurements of the system, and calculation of a single metric value to assess the stability of an observed site over any time period. The modeling framework presented here is just one of several viable approaches to addressing the question of stability and only robust statistical modeling and validation will allow us to form a complete picture of the natural world from a computational perspective.
Challenges and Assumptions for Stability Modeling

Understanding refugia areas, based on the climatic and topographic factors, is also driven by the availability (or lack thereof) of quality climate data. Global climate models have been constructed by research organizations, and especially The Intergovernmental Panel on Climate Change (IPCC) to simulate and measure a suite of variables such as temperature, precipitation, and atmospheric motion, in order to understand and anticipate historical and future climatic conditions. Methods to analyze climate systems are primarily based on the Coupled Model Intercomparison Project Phase 5 (CMIP5) database, formed for the 5th IPCC Assessment Report. From this model framework, numerous model interpretations have been developed, with differing biases and assumptions for different spatial and temporal scales. The modeling approach presented in this thesis project uses a single climate model in a historical context, but similar methodologies could be readily applied to future projections of climate as well. The suite of CMIP5 models each consider future climatic conditions, accounting for each emissions scenario, known as representative concentration pathways (RCP), which vary in severity and projection of emissions from RCP 2.6 (least severe) to RCP 8.5 (most severe; Wootten et al., 2014). Climate data can thus be used, depending on the RCP scenario, to determine trends and expectations for changes in temperature and precipitation, particularly in understanding the stability of these variables over time.

Identifying and classifying refugia are also limited by the availability and scale of the data and computational resources on the subject. Combining large-scale General Circulation Models with existing ground-based data forms the fundamental basis of this
study. Models of climate refugia should account for the specifics of a particular site’s relationship to its climate, including annual variability and landscape-scale characteristics (Reside et al., 2013). In future developments and refinements of this model, combining multiple modeling techniques and datasets will ultimately produce the greatest confidence in capturing the functionality of climate refugia on the landscape. Such comprehensive models should be presented, not only in a manner that forms a more complete picture of ecosystem structure, but also such that the results are readily accessible and interpretable to conservation planners, fellow researchers, and especially the public.

**Conservation Applications**

**Putting Climate Refugia into Practice**

Applying the results of modeled climate refugia is a relatively new practice within the conservation world and has yet to be adequately translated into practical action on the ground (NCCARF, 2015). While the use of topoclimate models is fairly new in the field of landscape ecology, the formulation of a framework to apply these models has been advancing. In their landmark paper, Morelli et al. 2016 outline the dynamic and challenging nature of identifying the relevant goals for spatial and temporal scale of refugia, analyzing the specific climatic features of the landscape in question, then integrating these areas into conservation prioritization, with a great need for constant monitoring and readjustment (Morelli et al, 2016). A priority aim of this master’s thesis is to make an impact on this process within the conservation management field by providing an example of the process of refugial site identification. From the results and
methodologies worked out here, I hope to strengthen the receptivity and connections of conservation organizations to using geospatial technologies in identifying priority resources and priority areas of key conservation value from a climate perspective. I maintain that refugia assessments and tracking can become a leading way to prioritize conservation action across complex landscapes in the face of changing climate. This is because they provide compelling targets for long-term ecosystem sustainability. The framework worked out here is one of the first efforts to use modeling techniques to highlight the lessons of climate refugia from a stability perspective. Being able to measure and project refugia provides added ecosystem response indicators as changes in the biosphere unfold (Schwartz, 2012). Using this approach to understand both climate refugia, and the importance of their connectivity, will allow planners and conservationists to ensure the support of a wider range of priority species, and the faster adaptation of their planning to changing global climate (Nuñez et al., 2013; Maher et al., 2017).

Researchers have multiple opportunities to collaborate on understanding the adverse effects of changing climatic factors on ecosystems and the adaptive capacity of climate-sensitive species over multiple temporal and spatial scales (Füssel and Klein, 2006; Young et al., 2011). One such application of further refugial research will be the assessment of the reduction in the vulnerability of sites through the natural buffering effects of refugia reducing exposure, as well as modeling the conservation value of such sites. A refugia protection approach to conservation prioritization will be an additional way to increase the value of protecting these potentially irreplaceable small and connected areas (Trombulak et al., 2008). Such analyses could consider modeled
response to climate change in current and future ranges, the interconnected nature of habitat and climate, and indirect factors such as sea level rise or natural and anthropogenic barriers to dispersal (Trombulak et al., 2008; Young et al., 2011). Modeling where these areas of greatest concern occur will likely lead to better prioritization of necessary landscape structures and the associated climate refugia.
CHAPTER TWO
RESEARCH DESIGN AND METHODS

Purpose of the Study

Constructing a model for predicting or validating the theoretical location of climate refugia, based on the relative stability of sites, allows for the identification and cataloging of areas of potential conservation prioritization in the Southern Appalachian region. The long-term development goal of this model based on the preliminary research presented here is to establish a well-accepted and useful methodology for categorizing the effects of the topographic complexity, species distributional shifts, and past and future disturbance regimes of this region on many Eastern North American species. In the next years of my research studies, I hope to continue contributing to the field’s growing understanding of how driving forces of refugia preserve diverse ecosystems.

Explanation of Research Design

I selected a series of methods to quantify the stability of sampled sites and determine the relationship between the measured climatic variables and the underlying topographic predictors. This analysis accounted for the variation across a recent period of recorded climatic history (1950 to 2005) and modeled the likelihood of refugial sites existing across the landscape. To accomplish this, climate data and topographic variables had to be selected and sampled in a usable and comparable format for statistical modeling. Once models had been created for the landscape, the topographic predictor variables could be spatially projected on the landscape (Figure 2.1).
Methodology Overview

**Multivariate Adaptive Constructed Analogs (MACA) Dataset**

- CanESM2 Global Climate Model

**Selected Variables** (daily):
- Temperature Max
- Temperature Min
- Precipitation Average
- Specific Humidity Average
- Surface Radiation Average

- Clipped and Projected to Study Area extent
- Converted into Raster format from .ncd

**Summarized Variables** (monthly):
- Maximum for each climate variable at monthly time step
- Mean

- Random points generated within gridded subsamples
- Climate data extracted to sampled points

**Correction applied to each Month:**
All values for each month of each variable were averaged, then subtracted from each value in that set to get the relative variation

- Permutations of subsets of climate variables generated

**Multivariate Control Chart** for each Site:
All values for each site used with a set of climate variables to construct control chart limits, then number of observations exceeding those limits recorded

**National Elevation Dataset (NED)**

- 1/3-arc-second resolution

**Calculated Variables**:
- Elevation
- Slope
- Aspect
- Topographic Position Index
- Topographic Ruggedness Index
- Roughness

**Summarized Variables** (focal analysis):
Topographic factors summarized to 150x resolution window to match climate data

- Topographic data extracted to sampled points
- Correlations calculated between variables

**Joined Climate and Topographic Data**:
Dataframe generated with each row as a site, XY coordinates, and associated topographic and control chart data

**Statistical Model Analysis and Testing**:
Generalized Linear Model produced, accounting for linear spatial correlation structure, and analyzed

**Final Output and Results**:
Statistical model and mapped outputs

Figure 2.1: Flowchart of Analysis (data inputs, processing, modeling, and output)
Summary of Methods

Study Area

This study focused on the central and southern Appalachian range of North America, specifically the Appalachian Forests (eastern portion of Level II Ecoregion 8.4 as delineated by the United States Environmental Protection Agency), including 13 states: Pennsylvania, Ohio, West Virginia, Maryland, Virginia, Kentucky, Tennessee, North Carolina, Georgia, and Alabama, as well as portions of New York, New Jersey, and South Carolina (Figure 2.2). This area was historically subject to glacial climatic conditions, which established the environmental baseline for the natural areas that the region is known for today. The wide range of heterogeneous habitat, topography, and species biodiversity has led to the region being considered the center of species richness and endemism for the eastern United States (Loehle, 2007). Three major eastern North American tree taxa have shown historical distribution throughout the region: spruces, including white spruce (*Picea glauca*), black spruce (*P. mariana*), and red spruce (*P. rubens*); eastern oaks; and eastern boreal pines (Morin and Unger, 1997). Evidence of historical climate refugia exists in the Appalachians, based on the ancient floristic distribution through mountain systems and associated soils and climates in the region, with the Ridge and Valley physiographical region of particular interest due to the potential for range expansion and migration (Loehle, 2007; Gonzales et al., 2008).
Figure 2.2: Map of the Study Area (eastern portion of Level II Ecoregion 8.4) (highlighted in dark green, designated 8.4) (United States Environmental Protection Agency, 2006)

**Climate Data**

The data used to construct the model was assembled from the Multivariate Adaptive Constructed Analogs (MACA) Dataset (v2. Livneh product) (Livneh et al., 2013), produced by the Northwest Climate Science Center and the University of Idaho. The MACA process downscaled the CanESM2 Global Climate Model, based on the Coupled Model Inter-Comparison Project 5 (CMIP5), and interpolated across the entire
contiguous United States. The method first produced a coarse bias correction to avoid stationarity in future model outputs, then formed daily constructed analogs based on patterns of similar climate measurements in 45-day windows across all sampled years. Reducing instances of stationarity allows for the mean and variance of the system to change over time and permits the statistical properties of the system to differ from the past and future (Nau, 2014). This data allowed for a more accurate downscaling process that reduced bias and increased compatibility across the datasets. The data sampled were provided at the 1/16-deg resolution on a daily temporal scale. The CanESM2 model was selected following an evaluation conducted specifically for the Southeast United States, which assessed the ability of the dataset to reproduce the observed climate trends of the 20th century in the region (Rupp, 2016). The MACA downscaling process preserves the dependencies between the variables and accounts for spatial patterns in the region, rather than relying solely on interpolation across the global extent (Rupp, 2016). The selected climate variables include maximum daily temperature (TasMax), minimum daily temperature (TasMin), average daily precipitation amount (Pr), average daily specific humidity (Huss), and average daily downward shortwave radiation (RSDS) (Table 2.1).

<table>
<thead>
<tr>
<th>Abbrev.</th>
<th>Description</th>
<th>Included in Final Model?</th>
</tr>
</thead>
<tbody>
<tr>
<td>TasMax</td>
<td>Maximum daily temperature near surface</td>
<td>Mean: No / Max: Yes / Min: Yes</td>
</tr>
<tr>
<td>TasMin</td>
<td>Minimum daily temperature near surface</td>
<td>Mean: Yes / Max: Yes / Min: No</td>
</tr>
<tr>
<td>Pr</td>
<td>Average daily precipitation at surface</td>
<td>Mean: No / Max: Yes / Min: No</td>
</tr>
<tr>
<td>Huss</td>
<td>Average daily specific humidity near surface</td>
<td>Mean: Yes / Max: Yes / Min: Yes</td>
</tr>
<tr>
<td>RSDS</td>
<td>Average daily downward shortwave radiation at surface</td>
<td>Mean: Yes / Max: Yes / Min: Yes</td>
</tr>
</tbody>
</table>

Table 2.1: Summary of Climate Variables included for final modeling: maximum daily temperature (TasMax), minimum daily temperature (TasMin), average daily precipitation amount (Pr), average daily specific humidity (Huss), and average daily downward shortwave radiation (RSDS) (Multivariate Adaptive Constructed Analogs)
**Landscape Topography Data**

Topographic information was derived from the National Elevation Dataset (NED) from the United States Geological Survey at a 1/3-arc-second resolution. From this elevation layer, several topographic variables were calculated, including the slope, aspect, topographic position index, topographic ruggedness index, and roughness value. The data were summarized to reflect the scale of the climate data using a focal analysis at a 150x resolution window, averaging each topographic variable over that extent. After determining which variables were correlated above a 0.9 level based on their Pearson correlation coefficient and removing contributing factors, the predictor variables tested in the model were elevation, slope, aspect, and the topographic position index (Table 2.2).

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
<th>Included in Final Model?</th>
</tr>
</thead>
<tbody>
<tr>
<td>NED</td>
<td>Elevation</td>
<td>Yes</td>
</tr>
<tr>
<td>Slope</td>
<td>Percent Rise value</td>
<td>Yes</td>
</tr>
<tr>
<td>Aspect</td>
<td>Degree of directional heading</td>
<td>Yes</td>
</tr>
<tr>
<td>TPI</td>
<td>Topographic position index on slope</td>
<td>Yes</td>
</tr>
<tr>
<td>TRI</td>
<td>Topographic ruggedness index amount</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>of elevation difference between</td>
<td></td>
</tr>
<tr>
<td></td>
<td>adjacent elevation cells</td>
<td></td>
</tr>
<tr>
<td>Roughness</td>
<td>Variation across a surface</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 2.2: Summary of Topographic Variable assembled for modeling: National Elevation Dataset (NED), Slope, Aspect, Topographic Position Index (TPI), Topographic Ruggedness Index (TRI), and Roughness (U.S. Geological Survey)
**Sampling and Processing**

The data were extracted from the native .ncd form into a GeoTIFF raster format, which separated each daily time-step into a series of individual layer files by variable name and date. Each layer was projected to the Geographic Coordinate System WGS 1984 and cropped to the study area extent, in order to reduce the necessary file storage and processing time. Additionally, the daily data would need to be reduced to monthly values to reduce processing load and reduce the effects of daily variation. Once the data were assembled into monthly stacks, each set of months were aggregated by summarizing by maximum, minimum, and mean for each climate variable, in order to capture a variety of statistical measures. These monthly stacks were then sampled by a set of randomly-generated points that had been defined by a gridded subsampling operation. Each sample point recorded the climate variable values for each of the three summary statistics on the monthly time-step, and the resulting data-frame was organized and sanitized for missing values.

Once the full climate variable samples were assembled, a month-based correction was applied to the dataset in order to compare monthly summaries across the entire year. Without such a correction, variation of non-extreme months would be underrepresented in the final control chart methodology. To calculate this for each value, the average value of each statistic was determined for each month of the year at each point, and this value was subtracted from the corresponding data, such that each resulting value was a difference from the mean of all months (See Appendix A for full code write-up).
Control Chart Stability Assessment

In order to determine a value for the relative stability of sites, the quality control chart statistical tool was utilized to measure a site’s climatic variation outside of calculated boundaries. While this technique is typically used to adjust a mechanical system to producing stable results, in this case, the calculation yields a measure of sites that exceed the bounds of normal variation in a measurable way. The control chart method generates statistical limits at three standard deviations from the average center line for each sampled site, and the number of observations outside of these limiting boundaries is recorded. This value indicates the metric for the total variability of a site’s climate over time and accounted for each site’s particular set of climatic conditions. The control chart measure is demonstrated here first in a univariate context (Figure 2.3), and then modified to run in a multivariate process, using the “mqcc” function in the “Quality Control Charts” (qcc) package in R Statistical Software (Montgomery, 2009; R Core Team, 2013). The multivariate quality control charts are constructed using the Hotelling $T^2$ statistic for each observation along the mean vector and the limits are based on the jointly constructed control regions of each variable. As noted in Table 2.2, a subset of the total permutations of the samples variables contributed to the final limiting boundaries. These variables were selected by running all permutations of combinations of eleven or more variables through the modeling framework and noting the best performing sets based on the relative AIC values. The best performing set included: the maximum and minimum daily maximum temperatures; maximum and mean daily minimum
temperatures; maximum daily precipitation; mean, maximum, and minimum humidity; and mean, maximum, and minimum downward shortwave radiation.

Figure 2.3: Example of Control Chart (univariate assessment of a single observed site) (R Statistical Software – qcc package). Along the x-axis are the monthly observations and the y-axis denotes the mean of maximum daily temperatures for each observation. Each red point above or below the upper and lower control bounds is beyond the limits established for each site, with the total count recorded as “Number beyond limits”.

**Statistical Model Creation**

Once the data were assembled and processed, generalized linear mixed models were produced using the PROC GLIMMIX procedure in the SAS statistical software (SAS Institute, 2013). The general linear mixed model allowed for the extension of a generalized linear model to incorporate normally distributed random effects of spatially-measured data. This method accounts for potential correlations between locations and the corresponding spatial structures. The resulting statistical output was reported and projected back onto the spatial dataset (Appendix B). A generalized least squares
methodology was also conducted in R, with model outputs producing similar results, however the SAS methodology was chosen for the aforementioned spatial considerations.

RESULTS

The analysis resulted in a likelihood model of relationships between the sampled climate-based stability value and the measured topographic parameters (Figure 2.4). In the model used to construct the final output, the generalized linear mixed models produced relatively robust relationships between the sampled climate values and elevation (AIC: 3086.2, pseudo-$R^2$: 0.3112674; Appendix B). The model suggests that greater climatic variability occurs at sites of lower elevation. The relationship between the predictor of elevation was statistically significant at an alpha level of 0.05, and nearly achieved a level of 0.01, while the other predictors did not result in a significant relationship. Therefore the highest likelihoods of stable climate refugial sites appears to have occurred along the most pronounced high elevation mountain regions of the region, corresponding to the relationships between climate and the topographic structures.

The greatest likelihood of refugia occurred along the Blue Ridge Mountains of Tennessee and northern Georgia, as well as along the Appalachian Plateau of West Virginia and eastern Kentucky. The lowest potential occurred along the western edge of the Piedmont through North Carolina and Virginia, up into central Pennsylvania. The northern tip of the Valley and Ridge landscape up into eastern Pennsylvania could also have a high likelihood of refugial sites.
Figure 2.4: A) Refugial Index projected across the Southern Appalachians based on the generalized linear mixed model comparing the variability of climate from 1950-2005 with sampled topographic features (darker color indicates a lower calculated index value, which corresponds with higher stability of sites, indicating the potential presence of climate refugia at this scale; B) Top 80% of model-based index value for demonstrating the most likely location of refugial sites (highlighted in red outlines); C) Top 50% of model-based index value for demonstrating all likely locations of refugial sites (ESRI – ArcMap GIS Software)
DISCUSSION

Understanding the Model Results

At this first stage of methodology the model output successfully demonstrated that certain factors such as elevation can be used in this type of model to define the most stable refugia areas within this region. Mountaintops and ridges appeared to be better captured by this modeling process. This stands in contrast to the generally held belief that low-elevation areas within mountainous areas have the best buffering potential for refugial sites. One reason might be that the summary procedure for measuring the landscape features averaging the topographic variation across each sampled pixel, at a larger scale than other prominent analyses on this subject. In doing so, the model could underrepresent features on the landscape such as coves and ridges, as these are too fine-scale to be picked up by this approach, or overestimate the stability of highly exposed mountaintops. While the topographic variables measured here likely accounted for broad heterogeneity of the landscape, there is less evidence of capturing cold air pooling or temperature inversions at this scale, which reduces the ability of the model to incorporate decoupling as a conceptual result (Dobrowski, 2011). In order to capture these relationships, we may investigate these topoclimate relationships at finer scales, as analyses on the scale of 30-meters can add up to 8 °C of variability over fine scale local temperature sensors (Ackerly et al., 2010). Additionally, the results undoubtedly include non-refugial areas, as the generalized nature of the model accounts for climate stability in any location, regardless of species-level factors (Ashcroft, 2010). More detailed topographic modeling may also allow for more complex statistical relationships to be
understood, particularly in testing the assumptions that these relationships are linear between the predictor variables and climatological response. Such linear relationships have been called into question based on experiments comparing land surface temperature with lapse rates (Oyler et al., 2016). The results presented here appear to be able to explain a limited number of factors involved in topoclimatic relationships; nevertheless, the foundations established by this methodology could be readily iterated upon for continuation of future research.

Justification for and Overview of Methodological Decisions and Their Limits

The focus of this project was primarily on the abiotic relationships between climate on the landscape and the influences of topography. In this way, the resulting models were both generalized and not specific to any one organism. The intention of such a model is to serve as a base-line for future research that may tailor this modeling framework to a set of taxa or habitats. The model was constructed using datasets validated for the Southern Appalachian region and was built using climate data from a single GCM. Further research could iterate on these testing methods to determine how other climate models may affect the outputs of this methodology. Additionally, a single scale was selected for the sampling of the climate and topographic data. The scale was both computationally feasible and comparable between the two datasets, though in future work, the effects of scale could be more explicitly modeled. Lastly, the outputs here should be considered to summarize the climatic relationships of the second half of the 20th century and may serve as a starting point for generating models of future climate
refugia. Such models would likely account for multiple climate projections and compare the outcomes of each scenario to prioritize the areas under greatest possible threat.

**Connections and Conclusions**

It is widely understood that geodiversity promotes biodiversity and provides the foundation of conserving valuable ecosystem structures and that geophysical variation, latitude, and elevation range were best predictors of species diversity (Cogbill and White, 1991; Anderson and Ferree, 2010; Hjort et al., 2015). The next set of relevant questions for conservation design is how do we measure topographic factors on the landscape in a meaningful way for defining which areas are of greatest prioritization interest. In addition to broadly defined refugial areas, how are these sites applicable to threatened taxa and can we predict habitat utilization and value over the long-term? Analyses of bioclimatic envelope models have demonstrated that incorporating topography into projections of species response to climate change shows an increase in accuracy of predictions for the species distributions, as well as demonstrating doubling of loss projections over climate-only models (Luoto and Keikkinen, 2008). Topographical heterogeneity has the potential to buffer against extinctions due to climate change, but these effects will need to be quantified across multiple gradients of habitat conditions and time periods to account for differences between organisms (Ashcroft, 2010). From a more generalized model like the one demonstrated in this study, a more refined species-specific model would likely reduce the predicted area of refugial presence. However, such models would need to explicitly examine errors of commission, as these may actually represent newly
established or potential functional niches outside of where the focal species currently inhabits (Araújo and Peterson, 2012). One may also ask if these refugial areas will continue to serve the same function for that species into the future and consider differing climate projections and the formation of non-analog climates (Veloz et al., 2012). Additionally, species on the landscape are undoubtedly influenced by a greater spectrum of factors than climate or topography alone, including spatial barriers, inter- and intra-species competitive behaviors, and anthropogenic pressures. We may also ask which species traits contribute to a reliance or preference for climate refugia, including specific climatic ranges, dispersal capabilities, high degrees of specialization, smaller population distributions, or limited phenotypic heterogeneity (Harrison and Noss, 2017). The relationships discussed here also assume that the niche stability of the species is maintained over time, such that the interactions expected between the species and its habitat remain the same (Keppel et al., 2012). The topography of the landscape is one piece of the numerous challenges in defining priority areas for the conservation of montane habitats and the next steps for research in this arena should be focused on constructing comprehensive models of refugial area that incorporate both the type of work demonstrated in this thesis with further analysis of taxon-specific distributional patterns.

Further Implications and Recommendations for Further Research

The research approach used in this project is a broad scale assessment of the Southern Appalachian region, to attempt to identify the prevalence of stable areas, based
on the topographic heterogeneity present across the landscape. While the use of control charts to identify stable sites is novel and still being explored, this type of multivariate technique has been shown here to capture at least a portion of the relationship between climate and the underlying topographic factors, in the category of a “refugia” but certainly not all of the presumed relationships of refugia. As a future study, by simultaneously accounting for a series of climate variables, and defining a statistically robust rationale for their variation outside of relative boundaries across the study area, we may be able to measure the expression of climate on the landscape on a more granular temporal spectrum. From a spatial perspective, such coarse-scale methods may be better at identifying the approximate locations of more fine-scale refugia and represent a broader categorization for prioritization metrics (Ashcroft, 2010).

The concept of climate refugia will continue to serve as an ideal for classifying critical habitat into the future, particularly as species congregate in areas of suitable environmental parameters (Loarie et al., 2009). With a greater concentration of climate-sensitive organisms in fewer locations comes greater inherent risk of impactful disturbance events, therefore protecting these areas is of vital importance to the long-term viability of montane ecosystems. Climate refugia may have greater likelihood of remaining stable sites for endemic species into the future; however, these sites still have the potential to experience significant shifts in climate patterns, which may be untenable for many species (Harrison and Noss, 2017). To ultimately implement models of climate refugia into conservation practice, both the process- and pattern-based methods of assessment must be integrated for the highest degree of confidence in the future
importance of key habitats on the landscape. Additionally, local knowledge and awareness of stable sites can be gathered and incorporated into the conservation design recommendations, both to increase the impact of such analyses on Appalachian communities, as well as communicate the importance of these areas for future generations. Through the identification of climate refugia primarily in the Southern Appalachian Mountains, the methodologies performed here may serve as a model for identifying stable sites for conservation practices in a variety of terrestrial landscapes for many generations to come.

Future assessments of climate refugia areas from a process-based perspective should consider incorporating more topographically-explicit datasets of predicted climatic patterns, such as TopoWx, which improves complex temperature trends, particularly for reversing the over-estimation of minimum temperatures in climate models. Temperature observations in mountainous regions have been shown to be artificially amplified by biases in the measurement systems, that can be corrected by modeling for these overestimations (Oyler, et al. 2015). The scale at which landscape characteristics are measured can also have a significant effect on the modeling results in defining refugial areas. Landform information is particularly sensitive to scaling, and future research should account for model outputs at a spectrum of scales to better quantify the relationship between the physiographic patterns, vegetation on the landscape, and organismal distribution (Theobald et al., 2015). At a finer scale, factors such as topographic convergence and cold air pooling could define specific topoclimatic predictors for identifying ideal habitat for many organisms of interest (Curtis et al.,
Additionally, our confidence in these model systems can be quantified by the degree to which the models have been iterated upon, accounting for a greater number of process-based factors, as well as using independent datasets at finer resolutions to validate the relationships delineated by the statistical outputs. The concepts discussed here are likely most strongly applicable in topographically complex systems, as the climate-topography relationships are likely not as strong in flatter topographies such as the coastal plain. However, we may use similar methods to those proposed here to determine the causal relationship between landscape features such as wetlands, forests, and water-bodies to the surrounding climatic conditions and pressures on species. Such broad models of refugia may not perfectly capture the climate refugia for all threatened organisms on the landscape; however, testing a wide range of spatial scales, climate scenarios, and observations of climate-sensitive species may lead to generally accepted paradigms for defining prioritization best practices from a topographic perspective.

Concurrently to a more refined methodology for understanding the influences of topographic parameters, the biological patterns in montane systems with regards to stable areas will be more comprehensively evaluated. Determining the maximum contraction of geographical ranges for climate-limited species will allow researchers to define refugial potential on a more specific basis, after which the refugial areas can be assessed for their carrying capacity and likelihood of supporting populations over a given time period into future climate scenarios (Stewart et al., 2009; Sutton et al., 2014). Organisms must also be able to match the pace of climatic change to persist in montane regions, which, especially for plants such as tree species, has been shown to have extreme migration
requirements as climates envelopes shift (Roberts and Hamann, 2016). The capacity of an organism to shift their distribution and persist in critical habitat requires significant biological knowledge and modeling specificity in a particular landscape. Modeling efforts to understand key refugial areas from a biological perspective may benefit from greater detail and also from a broader taxa perspective, combining the overlapping refugial ranges of multiple target species can create a more comprehensive prioritization map (Stewart et al., 2009; Loehle, 2011).

Through this research, I hope to highlight the lessons of climate refugia study to provide informative, responsive indicators to the relationship of climate to montane organisms, as changes in the biosphere unfold. This foundational work should serve as one of several possible inputs into a wider analysis of climate systems in the Southern Appalachians and will be followed up upon in future research endeavors. There is still much to be done to more fully understand how best to maintain a functional and sustainable ecosystem structure in the face of changing global pressures and climate refugia may serve as one important target for conservation design, particularly in topographically complex regions. While refugia as described here may not be the ultimate panacea for conserving threatened organisms, prioritizing these stable sites and the associated species has the potential to allow us to maintain the function of habitats to promote ecosystem persistence and conservation across a variety of landscapes in a changing global climate regime.
Appendix A

Final R Statistical Software Script

Data Processing and Analysis for Master's Thesis

Alec Nelson

June 7, 2017

Set Up Directories and Packages

```r
basedirectory <- "D:/DataAnalysis"
inputdata_path <- "D:/DataAnalysis/TestData"
countryshape_path <- "D:/DataAnalysis/App_Boundary_SHP"
outputdata_path <- "D:/DataAnalysis/TestResults5"
monthlydata_path <- "D:/DataAnalysis/MonthlyData3"
sampleddata_path <- "D:/DataAnalysis/TestSample"
SampledCSV_path <- "D:/DataAnalysis/SampledClimateData"
EnvirData_path <- "D:/DataAnalysis/EnvironmentData"
CombinIter_NoSeas_path <- "D:/DataAnalysis/MultiVarLimit_Iter_NO_SAdj"
CombinIter_SeasAdj_path <- "D:/DataAnalysis/MultiVarLimit_Iter_SeasonAdj"

setwd(inputdata_path)
list_of_packages <- c("raster","pracma","reshape","car","compute.es","effects","rgdal","fields","chron","ff","downloader","magrittr","maptools","GSIF","rgeos","ggplot2","multcomp","pastecs","data.table","MuMIn","ncdf4","sp","dismo","stringr","data.table","RCurl","rio","RNetCDF","parallel","qicharts","qcc","zoo","dplyr","purrr","plyr","geoR","geoRglm","MSQC","coda","MASS","relaimpo","arcgisbinding","lme4","glmm","nlme","arm","rms","rmarkdown")
new.packages <- list_of_packages[!(list_of_packages %in% installed.packages())[,"Package"]]
if(length(new.packages)){install.packages(new.packages)}

#Load all packages
lapply(list_of.packages, require, character.only = TRUE)
```
Process Climate Data to Convert, Clip, Project, and Summarize by Monthly Values

```r
for(f in 1:length(raster_file_list)){
    raster_file_name<-raster_file_list[f]
    pos = regexpr('_', raster_file_name)[1]
    var_name.f<-substr(raster_file_name, 1, (pos-1))
    raster_years<-substr(raster_file_name, nchar(raster_file_name)-23, nchar(raster_file_name)-15)
    outputpath_name.f<-paste0("D:/DataAnalysis/DailyData_MACA/Daily_",var_name.f,"_",raster_years)
    dir.create(outputpath_name.f)
    outputdata_path <- outputpath_name.f
    ###########################################################################
    setwd(inputdata_path)
    nc.i<-nc_open(as.character(raster_file_list[f]))
    var.i<-ncatt_get(nc.i, names(nc.i$var))$standard_name
    v3 <- ncatt_get(nc.i, names(nc.i$var))$standard_name
    v2 <- nc.i$var[[1]]
    varsize <- v2$varsize
    ndims <- v2$ndims
    nt <- varsize[ndims]
    nc_close(nc.i)
    print(paste0("Loaded parameters for:",as.character(raster_file_list[f])))

    for( i in 1:nt ) {
        setwd(inputdata_path)
        data.r<-raster( as.character(raster_file_list[f]),ymn = 25.125, ymx = 52.875, xmn = 235.375, xmx = 293,
                        ncdf=TRUE, varname=v3,lvar=3,level=1,band=i)
        timeval<-getZ(data.r)
        shift.raster <- raster::shift(data.r,-360)
        data.projected<-projectRaster( shift.raster , crs = proj4string( countryshape ),method = "bilinear" )
        data.proj.r<-raster::crop( data.projected , extent( countryshape ) )
        filename.i<-paste0(substr(raster_file_list[f],1,(nchar(raster_file_list[f])-24)),timeval,".tif")
        print(paste0("Saving file: ",filename.i))
        setwd(outputdata_path)
        rf <- writeRaster(data.proj.r, filename=filename.i, format="GTiff", overwrite=TRUE)
        setwd(inputdata_path)}
    ###########################################################################
```
# Stack each set of the raster layers
setwd(outputdata_path)
clipped_file_list <- list.files(pattern = "tif", all.files = FALSE, full.names = FALSE)

stack.g <- stack()
for(g in 1:length(clipped_file_list)){
    stack.g <- stack(stack.g, raster(clipped_file_list[g]))
    print(paste("Added file", clipped_file_list[g], "to stack"))
}

# Summarize data by monthly raster layers
stats.m <- c("Mean","Max","Min")
output.stat.months <- c("D:/DataAnalysis/MonthlyData_MACA/MonthlyData_Mean","D:/DataAnalysis/MonthlyData_MACA/MonthlyData_Max","D:/DataAnalysis/MonthlyData_MACA/MonthlyData_Min")

output.stat.months.f1 <- paste0(output.stat.months[1], "_", var_name.f, "_", raster_years)
dir.create(output.stat.months.f1, showWarnings = FALSE)
output.stat.months.f2 <- paste0(output.stat.months[2], "_", var_name.f, "_", raster_years)
dir.create(output.stat.months.f2, showWarnings = FALSE)
output.stat.months.f3 <- paste0(output.stat.months[3], "_", var_name.f, "_", raster_years)
dir.create(output.stat.months.f3, showWarnings = FALSE)

output.stat.months <- c(output.stat.months.f1, output.stat.months.f2, output.stat.months.f3)

stack.j <- stack.g
for(m in 1:length(stats.m)){
    month.prev <- 0
    year.prev <- 1950
    stack.month.i <- stack()
    for(k in 1:nlayers(stack.j)){
        Date.k <- 
        (substr(as.character(names(stack.j)[k]), nchar(names(stack.j)[k]) - 9, nchar(names(stack.j)[k])))
        Date.k <- gsub("[.]", ",", Date.k)
        Month.k <- as.numeric(format(as.Date(Date.k, origin = "1900-01-01"), "%m"))
        Month.char.k <- format(as.Date(Date.k, origin = "1900-01-01"), "%m")
        Year.k <- as.numeric(format(as.Date(Date.k, origin = "1900-01-01"), "%Y"))
        Year.char.k <- format(as.Date(Date.k, origin = "1900-01-01"), "%Y")
        if(Month.k > month.prev | Year.k > year.prev | k == nlayers(stack.j)){
            if(k > 1){
                
            } else {
                
            }
        }
    } else {
        
    }
}

} else {
    
}

}
if(stats.m[m]=="Mean"){
    stack.month.stat<- mean(stack.month.i, na.rm=TRUE)
    print(paste0("Calculated ", stats.m[m], " of Previous Month"))
} else if(stats.m[m]=="Max"){
    stack.month.stat<- max(stack.month.i, na.rm=TRUE)
    print(paste0("Calculated ", stats.m[m], " of Previous Month"))
} else if(stats.m[m]=="Min"){
    stack.month.stat<- min(stack.month.i, na.rm=TRUE)
    print(paste0("Calculated ", stats.m[m], " of Previous Month"))
}
else{
    print("ERROR: NO STATISTICAL FUNCTION PERFORMED!"
}
filename.k<- paste0(stats.m[m], ", ", (substr(as.character(names(stack.j)[k]), 1, nchar(names(stack.j)[k])-10)), year.prev.char.k, ", ", month.prev.char.k)
setwd(output.stat.months[m])
r.output <- writeRaster(stack.month.stat, filename=filename.k, format="GTiff", overwrite=TRUE)
setwd(outputdata_path)
}
stack.month.i<- stack()
stack.month.i<- stack(stack.month.i,stack.j[[k]])
print(paste("Began new stack w/", names(stack.j[[k]])," as first layer"))
} else{
    stack.month.i<- stack(stack.month.i,stack.j[[k]])
    print(paste("Added layer", names(stack.j[[k]])," to Month Stack"))
}
month.prev<- Month.k
month.prev.char.k<- Month.char.k
year.prev<- Year.k
year.prev.char.k<- Year.char.k

#########################################################################
#Stack all of the monthly raster layers and Sample data based on random points
stack.list<- list()
for(m in 1:length(stats.m)){
    setwd(output.stat.months[m])
    monthly_file_list <- list.files(pattern = ".tif", all.files = FALSE, full.names = FALSE)
    stack.m <- stack()
    for(g in 1:length(monthly_file_list)){

stack.m <- stack(stack.m, raster(monthly_file_list[g]))
print(paste("Added file", monthly_file_list[g], "to stack"))
assign(paste("stack.m", stats.m[m], sep=""), stack.m)
stack.add <- paste("stack.m", stats.m[m], sep="")
stack.list <- c(stack.list, eval(parse(text = stack.add)))
setwd(basedirectory))
samp_strat.points <- sample_points
samplepoints.df <- as.data.frame(coordinates(samp_strat.points))
samplepointnum <- c(1:nrow(samplepoints.df))
samplepoints.df <- cbind(samplepointnum, samplepoints.df)
for(x in 1:length(stack.list)){
  Samples.stack.mc <- data.frame()
  stack.mc <- stack.list[[x]]
  for(t in 1:nlayers(stack.mc)){
    layer.t <- stack.mc[[t]]
    layer.t.name <- names(layer.t)
    month.t <- substring(layer.t.name, nchar(layer.t.name) - 1, nchar(layer.t.name) - 6, nchar(layer.t.name) - 3)
    var.t <- substring(layer.t.name, 1, 4)
    extract.sample <- raster::extract(layer.t, samp_strat.points)
    extract.df <- as.data.frame(extract.sample)
    sample.names <- rep(layer.t.name, nrow(samplepoints.df))
    sample.month <- rep(month.t, nrow(samplepoints.df))
    sample.year <- rep(year.t, nrow(samplepoints.df))
    sample.coordinates.t <- cbind(sample.names, sample.year, sample.month, samplepoints.df, extract.df)
    Samples.stack.mc <- rbind(Samples.stack.mc, sample.coordinates.t)
  }
  print(paste("Sampled layer", layer.t.name, ", layer" , as.character(t), "out of", as.character(nlayers(stack.mc))))
  assign(paste("Samples.stack.mc.", stats.m[x], sep=""), Samples.stack.mc)

  colnames(Samples.stack.mc.Mean1) <- paste0(var_name.f, "_Mean")

  colnames(Samples.stack.mc.Max1) <- paste0(var_name.f, "_Max")

  colnames(Samples.stack.mc.Min1) <- paste0(var_name.f, "_Min")
  Samples.stack.total <- merge(Samples.stack.mc.Mean1, Samples.stack.mc.Max1, by=c("sample.year", "sa"
Samples.stack.total <-
merge(Samples.stack.total, Samples.stack.m.Min1, by=c("sample.year","sample.month","samplepointnum","coords.x1","coords.x2"), all=TRUE)
filename.f<-
paste0("D:/DataAnalysis/SampleStack_", var_name.f,"_",raster_years,".csv")
write.csv(Samples.stack.total, file=filename.f, row.names=FALSE)}

#########################################################################
#Combine groups of files into a single set of Samples
#########################################################################
setwd(SampledCSV_path)
csv_file_list <- list.files(pattern = ".csv" , all.files = FALSE , full.names = FALSE )
lastyear<- as.numeric(substr(csv_file_list[1],nchar(csv_file_list[1])-7,nchar(csv_file_list[1])-4))
filename.c<-
paste0()
Samples.1<- read.csv(csv_file_list[12])
Samples.2<- read.csv(csv_file_list[15])
Samples.3<- read.csv(csv_file_list[9])

Samples.TasTotal <-
merge(Samples.1,Samples.2,by=c("sample.year","sample.month","samplepointnum","coords.x1","coords.x2"), all=TRUE)
Samples.TasTotal <-
merge(Samples.TasTotal,Samples.3,by=c("sample.year","sample.month","samplepointnum","coords.x1","coords.x2"), all=TRUE)

Samples.1950_1969<- Samples.TasTotal
Samples.1970_1989<- Samples.TasTotal
Samples.1990_2005<- Samples.TasTotal

Samples.TasTotal<-

newdata <-
Samples.TasTotal[order(Samples.TasTotal$sample.year,Samples.TasTotal$sample.month,Samples.TasTotal$samplepointnum),]
write.csv(newdata, file="D:/DataAnalysis/SampleStack_ClimateTotal.csv", row.names=FALSE)

Apply a Monthly Correction to each Climate Variable
setwd(basedirectory)
SampleTotal<- read.csv("SampleStack_ClimateTotal.csv")
YearMonth<- as.yearmon(paste(SampleTotal$sample.year[1],SampleTotal$sample.month[1], sep="-"))
for(y in 2:nrow(SampleTotal)){
  YearMonth.y<- as.yearmon(paste(SampleTotal$sample.year[y],SampleTotal$sample.month[y], sep="-"))
  YearMonth<- c(YearMonth, YearMonth.y)
  percent<- (y/nrow(SampleTotal))*100
  print(paste("Added YearMonth for ", YearMonth.y, ", row", as.character(y), " out of ", as.character(nrow(SampleTotal)), "( ", percent, ") " ))}
SampleTotal.y<- cbind(SampleTotal, YearMonth)
row.ha.na<- apply(SampleTotal.y, 1, function(x){any(is.na(x))})
sum(row.ha.na)
SampleTotal.final<- SampleTotal.y[!row.ha.na,]

######################################################################
# Subtract Mean by Month for each point
setwd(basedirectory)
SampleTotal.final<- read.csv("SampleStack_ClimateTotal_Final.csv")
SampleTotal.final.pqrs<- SampleTotal.final
samplepointnum.unique<- unique(SampleTotal.final$samplepointnum)
samplepoint.month.unique<- unique(SampleTotal.final$sample.month)
vars.final<- names(SampleTotal.final)[6:20]
col.nums<- c(6:20)
for(p in 1:length(samplepointnum.unique)){
  sample.point.p<- SampleTotal.final[SampleTotal.final$samplepointnum == samplepointnum.unique[p],]
  for(q in 1:length(samplepoint.month.unique)){
    sample.month.p<- sample.point.p[sample.point.p$sample.month == samplepoint.month.unique[q],]
    for(r in 1:length(vars.final)){
      sample.mean.var<- mean(sample.month.p[, vars.final[[r]]])
      test.sample.rows<- which(SampleTotal.final$samplepointnum==samplepointnum.unique[p] & SampleTotal.final$sample.month == samplepoint.month.unique[q])
      for(s in 1:length(test.sample.rows)){
        SampleTotal.final.pqrs[test.sample.rows[s], vars.final[[r]]]<- (SampleTotal.final[s, vars.final[[r]]] - sample.mean.var)
      }
      print(paste("Adjusted values for month", q, "(Sample Point ", samplepointnum.unique[p], " ) "))
    }
  }
}
percent<- (p/length(samplepointnum.unique))*100
Test Univariate Control Chart Analysis on each Climate Variable

```r
setwd(basedirectory)
SampleTotal.SeasonAdj <- read.csv("SampleTotal_SeasonAdj.csv")
SampleTotal.final <- SampleTotal.SeasonAdj

# xbar samples
xbar.samples <- data.frame(matrix(ncol=31, nrow=length(unique(SampleTotal.final$samplepointnum))))
colnames(xbar.samples) <- c("samplepointnum", "tasmax_Mean.v.limits", "tasmax_Mean.v.runs", "tasmax_Max.v.limits", "tasmax_Max.v.runs", "tasmin_Mean.v.limits", "tasmin_Mean.v.runs", "tasmin_Max.v.limits", "tasmin_Max.v.runs", "huss_Mean.v.limits", "huss_Mean.v.runs", "huss_Max.v.limits", "huss_Max.v.runs", "pr_Mean.v.limits", "pr_Mean.v.runs", "pr_Max.v.limits", "pr_Max.v.runs", "pr_Min.v.limits", "pr_Min.v.runs", "rsds_Mean.v.limits", "rsds_Mean.v.runs", "rsds_Max.v.limits", "rsds_Max.v.runs")
samplepointnum.unique <- unique(SampleTotal.final$samplepointnum)
vars.final <- names(SampleTotal.final)[6:20]

for(x in 1:length(samplepointnum.unique)){
  test.sample <- SampleTotal.final[SampleTotal.final$samplepointnum == samplepointnum.unique[x],]
  xbar.samples[x,1] <- samplepointnum.unique[x]

  for(z in 1:length(vars.final)){
    z.var <- z+5
    xbar_chart1 <- qcc(data=test.sample[z.var], type="xbar.one", plot=FALSE, digits=5)
    sample.violation.limits <- length(xbar_chart1$violations$beyond.limits)
    sample.violation.runs <- length(xbar_chart1$violations$violating.runs)
    xbar.samples[x,(z*2)] <- sample.violation.limits
    xbar.samples[x,(z*2+1)] <- sample.violation.runs
    percent <- (x/length(unique(SampleTotal.final$samplepointnum)))*100
    print(paste("Tested Limits for", samplepointnum.unique[x], "samplenum ", ",", percent,"%") )
  }
}
```
Perform Multivariate Control Chart Analysis with all Climate Variables

```r
setwd(basedirectory)
SampleTotal.final <- read.csv("SampleStack_ClimateTotal_Final.csv")
SampleTotal.SeasonAdj <- read.csv("SampleTotal_SeasonAdj.csv")
# SampleTotal.final <- SampleTotal.SeasonAdj
SampleTotal.final$pr_Min <- NULL

mqcc.samples <- data.frame(matrix(ncol=2, nrow=length(unique(SampleTotal.final$samplepointnum))))
samplepointnum.unique <- unique(SampleTotal.final$samplepointnum)
vars.final <- names(SampleTotal.final)[6:19]

for(x in 1:length(samplepointnum.unique)){
  test.sample <- SampleTotal.final[SampleTotal.final$samplepointnum == samplepointnum.unique[x],]
  mqcc.samples[x,1] <- samplepointnum.unique[x]
  test.sample.x <- test.sample[,6:19]
  T2_single_chart1 <- mqcc(test.sample.x, type="T2.single", limits=TRUE, pred.limits = FALSE)
  sample.violation.limits <- length(T2_single_chart1$violations$behind.limits)
  mqcc.samples[x,2] <- sample.violation.limits
  percent <- (x/length(unique(SampleTotal.final$samplepointnum))) * 100
  print(paste("Tested Limits for", samplepointnum.unique[x], "samplenum ", ",", percent,"%" )))
}

names(mqcc.samples)[1] <- "samplepointnum"
names(mqcc.samples)[2] <- "multivar.v.limits"
SamplePointNums <- SampleTotal.final[3:5]
table.d <- SamplePointNums[1:length(unique(SamplePointNums$samplepointnum)),]
MQCC_SampleTotal <- merge(table.d, mqcc.samples, by="samplepointnum")
```

Import and process Topographic Raster Data

```r
setwd(basedirectory)
SampleTotal.final <- read.csv("SampleStack_ClimateTotal_Final.csv")
setwd(EnvirData_path)
NED.raster <- raster("AppLCC_ned_30_proj.tif")
```
#Apply the terrain functions to the DEM dataset

Slope.NED <- terrain(NED.raster, opt='slope', unit='degrees', neighbors=8, filename = "Slope_ned_proj.tif")
Aspect.NED <- terrain(NED.raster, opt='aspect', unit='degrees', neighbors=8, filename = "Aspect_ned_proj.tif")
TPI.NED <- terrain(NED.raster, opt='TPI', filename = "TPI_ned_proj.tif")
TRI.NED <- terrain(NED.raster, opt='TRI', filename = "TRI_ned_proj.tif")
Roughness.NED <- terrain(NED.raster, opt='roughness', filename = "Roughness_ned_proj.tif")

Slope.raster <- raster("Slope_ned_proj.tif")
Aspect.raster <- raster("Aspect_ned_proj.tif")
TPI.raster <- raster("TPI_ned_proj.tif")
TRI.raster <- raster("TRI_ned_proj.tif")
Roughness.raster <- raster("Roughness_ned_proj.tif")

samp_strat.points <- sample_points
samplepoints.df <- as.data.frame(coordinates(samp_strat.points))
samplepointnum <- c(1:nrow(samplepoints.df))
samplepoints.df <- cbind(samplepointnum, samplepoints.df)
NED.extract <- raster::extract(NED.raster, samp_strat.points)
Slope.extract <- raster::extract(Slope.raster, samp_strat.points)
Aspect.extract <- raster::extract(Aspect.raster, samp_strat.points)
TPI.extract <- raster::extract(TPI.raster, samp_strat.points)
TRI.extract <- raster::extract(TRI.raster, samp_strat.points)
Roughness.extract <- raster::extract(Roughness.raster, samp_strat.points)

setwd(EnvirData_path)
NEDfocal <- raster("NEDFocal_150x150_mean.tif")
Slopefocal <- raster("SlopeFocal2_150x150_mean.tif")
Aspectfocal <- raster("AspectFocal_150x150_mean.tif")
TPIfocal <- raster("focaltpi_150")
TRIfocal <- raster("focaltri_150")
Roughnessfocal <- raster("focalrou_150")

samp_strat.points <- sample_points
samplepoints.df <- as.data.frame(coordinates(samp_strat.points))
samplepointnum <- c(1:nrow(samplepoints.df))
samplepoints.df <- cbind(samplepointnum, samplepoints.df)
NED.extract <- raster::extract(NEDfocal, samp_strat.points)
Slope.extract <- raster::extract(Slopefocal, samp_strat.points)
Aspect.extract <- raster::extract(Aspectfocal, samp_strat.points)
TPI.extract <- raster::extract(TPIfocal, samp_strat.points)
TRI.extract<-raster::extract(TRI focal, samp_strat.points)
Roughness.extract<-raster::extract(Roughness focal, samp_strat.points)
NED.extract.df<-as.data.frame(NED.extract)
Slope.extract.df<-as.data.frame(Slope.extract)
Aspect.extract.df<-as.data.frame(Aspect.extract)
TPI.extract.df<-as.data.frame(TPI.extract)
TRI.extract.df<-as.data.frame(TRI.extract)
Roughness.extract.df<-as.data.frame(Roughness.extract)
NED.samples<-cbind(samplepoints.df,NED.extract.df,Slope.extract.df,Aspect.extract.df,TPI.extract.df,TRI.extract.df,Roughness.extract.df)

setwd(basedirectory)
MQCC_SampleTotal<-read.csv("MultiQCC_SampleTotal.csv")
MQCC_SampleTotal<--read.csv("MultiQCC_SampleTotal_SeasonAdj.csv")
MQCC_SampleTotal1<-merge(NED.samples,MQCC_SampleTotal,by=c("samplepointnum","coords.x1","coords.x2"))
MQCC_SampleTotal1<-MQCC_SampleTotal1[order(MQCC_SampleTotal1[,1]),]

names(MQCC_SampleTotal1)[2]<-"x1"
names(MQCC_SampleTotal1)[3]<-"x2"
names(MQCC_SampleTotal1)[4]<-"NEDextract"
names(MQCC_SampleTotal1)[5]<-"Slopeextract"
names(MQCC_SampleTotal1)[6]<-"Aspectextract"
names(MQCC_SampleTotal1)[7]<-"TPIextract"
names(MQCC_SampleTotal1)[8]<-"TRIextract"
names(MQCC_SampleTotal1)[9]<-"Roughnessextract"
names(MQCC_SampleTotal1)[10]<-"multivarlimits"

PointNums<-read.csv("AppLCC_Points.csv")
App.points<-PointNums[,1]
selectedRows <- (MQCC_SampleTotal1$samplepointnum %in% App.points)
MQCC_SampleTotal.App <- MQCC_SampleTotal1[selectedRows,]
MQCC_SampleTotal.App<-MQCC_SampleTotal.App[complete.cases(MQCC_SampleTotal.App),]

Generate Permutations of Climate Variables and Test Statistical Models

setwd(basedirectory)
MQCC_SampleTotal<-read.csv("Topo_MQCC_Focal_SampleTotal_SeasonAdj.csv")
var.col<-ncol(MQCC_SampleTotal)
setwd(basedirectory)
SampleTotal.final<-read.csv("SampleStack_ClimateTotal_Final.csv")
SampleTotal.SeasonAdj<-read.csv("SampleTotal_SeasonAdj.csv")
SampleTotal.final<-SampleTotal.SeasonAdj
SampleTotal.final$pr_Mi< - NULL

vars.total.all<-names(SampleTotal.final)[6:19]

###########
iter.comb<-c(11:14)
samplepointnum.unique<-unique(SampleTotal.final$samplepointnum)
SamplePointNums<-SampleTotal.final[3:5]
table.d<- SamplePointNums[1:length(unique(SamplePointNums$samplepointnum)),]
PointNums<-read.csv("AppLCC_Points.csv")
App.points<-PointNums[,1]

for(i in 1:length(iter.comb)){
  vars.list<-combn(names(SampleTotal.final)[6:19],iter.comb[i])
    for(c in 1:ncol(vars.list)){
      mqcc.samples<- data.frame(matrix(ncol=2,nrow=length(unique(SampleTotal.final$samplepointnum))))
      vars.final<-vars.list[,c]
        for(x in 1:length(samplepointnum.unique)){
          test.sample<-SampleTotal.final[SampleTotal.final$samplepointnum
== samplepointnum.unique[x],]
          mqcc.samples[x,1]<-samplepointnum.unique[x]
          test.sample.x<-test.sample[,vars.final]
          T2_single_chart1<-mqcc(test.sample.x,type="T2.single",limits=TRUE, pred.limits = FALSE)
          sample.violation.limits<-length(T2_single_chart1$violations$beyond.limits)
          mqcc.samples[x,2]<-sample.violation.limits
          names(mqcc.samples)[1]<-"samplepointnum"
          names(mqcc.samples)[2]<-"multivarvlimits"
          MQCC_SampleTotal <- 
          merge(table.d,mqcc.samples,by="samplepointnum")
          selectedRows <- (MQCC_SampleTotal$samplepointnum %in% App.points)
          MQCC_SampleTotal.App <- MQCC_SampleTotal[selectedRows,]
          MQCC_SampleTotal.App<-MQCC_SampleTotal.App[complete.cases(MQCC_SampleTotal.App),]
          index.nums<-which(vars.total.all %in% vars.final)
          index.coll<-paste(index.nums,collapse="~")
          filename.f<-paste0("D:/DataAnalysis/MultiVarLimit_Iter_SeasonAdj/MQCC_Limits_",index.coll,".csv")
          write.csv(MQCC_SampleTotal.App,file=filename.f,row.names=FALSE)
          print(paste("Measured Limits for",c," iteration out of ",ncol(vars.list),"(Combination Length",iter.comb[i],")" ))

}
```r
print(paste("Saved to:", filename.f)))

setwd(basedirectory)
MQCC_SampleTotal<- read.csv("Topo_MQCC_Focal_SampleTotal_SeasonAdj.csv")
NED.samples<-MQCC_SampleTotal[,,-10]
setwd(CombinIter_SeasAdj_path)
csv_file_list <- list.files(pattern = ".csv", all.files = FALSE, full.names = FALSE)
Model.compare.AIC<- data.frame(matrix(ncol=5,nrow=length(csv_file_list)))
names(Model.compare.AIC)<- c("Model.num","Model.vars","Model1.AIC","Model2.AIC","Model3.AIC")

for(i in 1:length(csv_file_list)){
  mqcc.samples<-read.csv(csv_file_list[i])
  names(mqcc.samples)[2]<-"x1"
  names(mqcc.samples)[3]<-"x2"
  MQCC_SampleTotal.merge <- merge(NED.samples,mqcc.samples,by=c("samplepointnum","x1","x2"))
  MQCC_SampleTotal.merge[order(MQCC_SampleTotal.merge[,1]),]
  Model.AIC.1<-9999
  Model.AIC.2<-9999
  Model.AIC.3<-9999
  Model.base.1<- NA
  Model.corAR1.2<- NA
  Model.corLIN.3<- NA

  if(min(MQCC_SampleTotal.merge$multivarvlimits)>0){
    tryCatch(
      Model.base.1<-gls(multivarvlimits ~ NEDextract + Slopeextract + Aspectextract + TPIextract,
      data = MQCC_SampleTotal.merge,method="REML"),error=function(e) {NA})
    tryCatch(
      Model.corAR1.2<-gls(multivarvlimits ~ NEDextract + Slopeextract + Aspectextract + TPIextract, correlation=corAR1(form=~1),
      data = MQCC_SampleTotal.merge,method="REML"),error=function(e) {NA})
    tryCatch(
      Model.corLIN.3<-gls(multivarvlimits ~ NEDextract + Slopeextract + Aspectextract + TPIextract, correlation=corLin(form=~ x1 + x2),
      data = MQCC_SampleTotal.merge,method="REML"),error=function(e) {NA})

    Model.AIC.1<-tryCatch(AICc(Model.base.1),error=function(e) {print(paste0("Error in Model 1 of iteration ",i));NA})
    Model.AIC.2<-tryCatch(AICc(Model.corAR1.2),error=function(e) {print(paste0("Error in Model 2 of iteration ",i));NA})
    Model.AIC.3<-tryCatch(AICc(Model.corLIN.3),error=function(e) {print(paste0("Error in Model 3 of iteration ",i));NA})
  }
}
```
```r
{print(paste0("Error in Model 2 of iteration ",i));NA})
Model.AIC.3<-tryCatch(AICc(Model.corLIN.3),error=function(e)
{print(paste0("Error in Model 3 of iteration ",i));NA})
if(!is.na(Model.AIC.1)){Model.compare.AIC[i,3]<-AICc(Model.base.1)}
if(!is.na(Model.AIC.2)){Model.compare.AIC[i,4]<-AICc(Model.corAR1.2)}
if(!is.na(Model.AIC.3)){Model.compare.AIC[i,5]<-AICc(Model.corLIN.3)}
filename.f<-csv_file_list[i]
col.nums.char<-substr(strsplit(filename.f,"_")[[1]][3],1,nchar(strsplit(filename.f,"_")[[1]][3])-4)
Model.compare.AIC[i,1]<-i
Model.compare.AIC[i,2]<-col.nums.char
print(paste("Models tested for",i,"iteration out of",length(csv_file_list)))
print(Model.compare.AIC[i,])
Model.compare.AIC<-read.csv("D:/DataAnalysis/Model.compare.AIC_SeasonAdj.csv")
Model.compare.AIC[which.min(Model.compare.AIC$Model1.AIC),]
Model.compare.AIC[which.min(Model.compare.AIC$Model2.AIC),]
Model.compare.AIC[which.min(Model.compare.AIC$Model3.AIC),]
Model.compare.AIC.sort<-Model.compare.AIC[order(Model.compare.AIC$Model3.AIC),]
col.nums.unique<-unique(na.omit(as.numeric(unlist(strsplit(unlist(Model.compare.AIC[420,2]),"[^0-9+]"))))))
vars.result<-vars.total.all[col.nums.unique]

mqcc.samples<-read.csv(csv_file_list[420])
names(mqcc.samples)[2]<-"x1"
names(mqcc.samples)[3]<-"x2"
MQCC_SampleTotal.merge <- merge(NED.samples,mqcc.samples,by=c("samplepointnum","x1","x2"))
MQCC_SampleTotal.merge[order(MQCC_SampleTotal.merge[,1]),]
Model.corAR1.2<-gls(multivarvlimits ~ NEDextract + Slopeextract + Aspectextract + TPIextract, correlation=corAR1(form=~1),
data = MQCC_SampleTotal.merge,method="REML")
Model.corLIN.3<-gls(multivarvlimits ~ NEDextract + Slopeextract + Aspectextract + TPIextract, correlation=corlin(form=~ x1 + x2),
data = MQCC_SampleTotal.merge,method="REML",verbose = TRUE)
summary(Model.corAR1.2)
summary(Model.corLIN.3)
```
Appendix B

Generalized Linear Mixed Model Statistical Output

Figure B-1: Output of Linear Mixed Model from SAS Statistical Software

<table>
<thead>
<tr>
<th>Fit Statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-2 Res Log Likelihood</td>
<td>3080.2</td>
</tr>
<tr>
<td>AIC (Smaller is Better)</td>
<td>3086.2</td>
</tr>
<tr>
<td>AICC (Smaller is Better)</td>
<td>3086.3</td>
</tr>
<tr>
<td>BIC (Smaller is Better)</td>
<td>3100.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PARMS Model Likelihood Ratio Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>DF</td>
</tr>
<tr>
<td>----</td>
</tr>
<tr>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Solution for Fixed Effects</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect</td>
<td>Estimate</td>
<td>Standard Error</td>
<td>DF</td>
<td>t Value</td>
<td>Pr &gt;</td>
</tr>
<tr>
<td>Intercept</td>
<td>12.8960</td>
<td>0.6924</td>
<td>62.4</td>
<td>18.62</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>NEDextract</td>
<td>-0.00095</td>
<td>0.000445</td>
<td>794</td>
<td>-2.14</td>
<td>0.0324</td>
</tr>
<tr>
<td>Slopeextract</td>
<td>0.02517</td>
<td>0.01545</td>
<td>905</td>
<td>1.63</td>
<td>0.1037</td>
</tr>
<tr>
<td>Aspectextract</td>
<td>0.002394</td>
<td>0.003093</td>
<td>779</td>
<td>0.77</td>
<td>0.4391</td>
</tr>
<tr>
<td>TPlextract</td>
<td>0.7733</td>
<td>2.0235</td>
<td>853</td>
<td>0.38</td>
<td>0.7024</td>
</tr>
</tbody>
</table>
REFERENCES


