HABITAT SUITABILITY MODELLING OF EASTERN TURKEYBEARD (*XEROPHYLLUM ASPHODELOIDES*) IN THE TALLADEGA NATIONAL FOREST

by

VOLKAN BEKTAS

(Under the Direction of PETE BETTINGER)

ABSTRACT

Habitat modeling methods are important for determining suitable conditions for species in order to manage natural areas sustainably. This study focused on the rare and threatened plant species eastern turkeybeard (*Xerophyllum asphodeloides*) in the Talladega National Forest in Alabama, United States. An ensemble suitable habitat map was developed using four different modeling methods (MaxEnt, Generalized Linear Model, Generalized Additive Model, and Random Forest). The models were successfully applied, and suitability maps were generated. This research suggested including more occurrence points on the modeling process, but also described the development and application of a suitable habitat map for the Talladega National Forest. This model can be used to inform managers on appropriate natural resource practices to maintain and restore sustainable forest management regimes with a conservation perspective. The knowledge gained from this research can be used to establish and implement habitat suitability strategies across the Talladega National Forest.

INDEX WORDS: Generalized Additive Model, Generalized Linear Model, Habitat Suitability Modelling, MaxEnt, Random Forest, Talladega National Forest, *Xerophyllum asphodeloides*.

HABITAT SUITABILITY MODELLING OF EASTERN TURKEYBEARD (*XEROPHLLUM ASPHODELOIDES*) IN THE TALLADEGA NATIONAL FOREST

by

VOLKAN BEKTAS

BS, Karadeniz Technical University, Turkey, 2013

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2021

HABITAT SUITABILITY MODELLING OF EASTERN TURKEYBEARD (*XEROPHYLLUM ASPHODELOIDES*) IN THE TALLADEGA NATIONAL FOREST

by

VOLKAN BEKTAS

Major Professor:	Pete Bettinger
Committee:	Nate Nibbelink
	Jacek Siry

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
August 2021

DEDICATION

To my lovely grandparents whom

I miss every day,

and I lost during this journey:

Fikriye Bektaş,

Züfer Bektaş,

Hüseyin Öztürk,

Fatma Öztürk.

I will always remember their endless love.

ACKNOWLEDGEMENTS

First of all, I would like to start by thanking my advisor, Dr. Pete Bettinger. I am grateful to him for always supporting, guiding, and encouraging me in all aspects throughout my studies. His friendly, always listening, supportive, and open-minded personality have always given me strength. I will always be grateful to him for having had the opportunity to work with him and for not being able to complete this work without his support. I will always remember all of his scientific advice and his “Science is hard” motto. This research would not have been possible without his endless help.

I would like to thank my committee members Dr. Nate Nibbelink and Dr. Jacek Siry for their support and their invaluable advice. They were always ready to help me every time, and their invaluable comments and advice helped my research a lot. I appreciate all their support, and I am glad to work with them.

I would also want to thank Krista Merry for her all advice, time, and endless help during my research. I am grateful to Jonathan Stober for his support and helps during this research. I would like to also give special thanks to Dr. Yanshu Li who gave me the opportunity to work as a research assistant at Warnell School of Forestry and Natural Resources. I am glad to get a chance to work with her, and I will always appreciate all her support and her very helpful and kind personality. I would also want to thank Dr. Cristian Montes, Dr. Roger Lowe, and all members of the Warnell School of Forestry and Natural Resources faculty who help me during my research. I have gained a lot of learning experience, and their support and friendship encouraged me during my research.

I would like to give special thanks to my supportive and helpful friends Dr. Shingo Obata and Katrina Henn who gave me the opportunity to have a great time during my research. I also want to thank Leslie Boby who allowed me to work with amazing projects. I am also glad to meet with Steven Weaver who left us. I will always remember his amazing personality. To my lab members Manlin, Taeyoon, Rocio, John, and Joshua, thanks for all the amazing time shared with me during my process.

I would like to say thanks to Dr. Oguz Kurdoglu and Dr. Hilmi Dogan for their encouragement and support to complete this journey. I would like to thank all my friends for their support, encouragement, and understanding during my journey from far away. To Oncu, Yucel, Murat Can, Ercan, Tugba, thanks for all their understanding and support. I would like to also thank people who support, helped me and gave me the chance to have good company here: Yasin, Ayse, Alp, Santiago, Damla, Nawa, Halil, Sehri, Laura, Sercan, Aytug, Wezzdy, Ilhan, Hatice, Pietro, Courtney, Thomas, John, and all the people who I met in the USA.

I would like to thank my lovely country Turkey and the Turkish Ministry of Education for the sponsorship and support for this master's program.

The last and biggest thanks are to my lovely family, who believed in me and supported me every time. I would like to appreciate their patient and endless love during my journey. Special thanks to my father Eyup Bektas, mother Turkan Bektas, my sister Yazgi Bektas Ozdemir, her husband Sedat Ozdemir, my sweet nephew Mustafa Cagan Ozdemir and my lovely wife Ekin Bektas. Without their support and encouragement, I would not be able to complete this journey. I will always remember their endless support and their unconditional love. I will always be glad to be a part of them.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS v

LIST OF TABLES vii

LIST OF FIGURES viii

CHAPTER

1 INTRODUCTION 1

2 METHODS 8

2.1. Study Area 8

2.2. Occurrence Data 10

2.3. Environmental Variables 12

2.4. Habitat Suitability Modeling 13

2.5. Model Evaluation 18

3 RESULTS 20

4 DISCUSSION 41

5 CONCLUSION 46

REFERENCES 48

APPENDICES

A Detailed Explanation of Bioclimatic Variables 59

B The Correlation Matrix for All Environmental Variables 60

LIST OF TABLES

Page

Table 1. VIFs of the 12 variables that remain after eliminating some based on Pearson's *r* and VIF scores. 20

Table 2. Coefficients Table for Generalized Linear Model. 25

Table 3. Coefficients Table for Generalized Additive Model. 28

Table 4. AUC Scores of All Modelling Methods. 34

Table 5. Suitable areas based on the models for the turkeybeard species in Talladega National Forest. 35

LIST OF FIGURES

Page

Figure 1.Talladega National Forest and turkeybeard occurrence locations in Talladega National Forest. (Basemaps sources: ESRI) 10

Figure 2. *Xerophyllum asphodeloides* (turkeybeard) in Talladega National Forest (Taken on 12.08.2020) 12

Figure 3. The correlation matrix for remaining variables whose correlations are under 0.7.......... 21

Figure 4. Contribution of the environmental variables in MaxEnt. 22

Figure 5. MaxEnt prediction map based on raw values. 23

Figure 6. MaxEnt prediction map based on presence and absence. 24

Figure 7. Generalized Linear Model prediction map based on raw values. 26

Figure 8. Generalized Linear Model prediction map based on presence and absence. 27

Figure 9. Generalized Additive Model prediction map based on raw values. 29

Figure 10. Generalized Additive Model prediction map based on presence and absence. 30

Figure 11. Contribution of the environmental variables in Random Forest. 31

Figure 12. Random Forest prediction map based on raw values. 32

Figure 13. Random Forest prediction map based on presence and absence. 33

Figure 14. *Xerophyllum asphodeloides* (turkeybeard) habitat suitability map based on approach 1. (Basemap source: ESRI) 36

Figure 15. *Xerophyllum asphodeloides* (turkeybeard) habitat suitability map based on approach 2. (Basemap source: ESRI) 37

Figure 16. *Xerophyllum asphodeloides* (turkeybeard) habitat suitability map for the entire forest. (Basemap source: ESRI) 38

Figure 17. Accuracy map between already known turkeybeard locations and predictions. 40

CHAPTER 1

INTRODUCTION

Turkeybeard (*Xerophyllum asphodeloides*) is a perennial understory forest herb that grows in discreet mountain populations in the eastern United States, from Virginia to Alabama. The plant occurs naturally in forested areas of these states as well in the Pine Barrens of Southern New Jersey and thus is found in 10 states (Nuttall, 1818; Fernald, 1950; Bourg et al., 2005; Bourg et al., 2015). Turkeybeard is a native plant for North America, and it is classified as a rare and threatened plant species in some states and at the federal level (Keener et al., 2020; United States Department of Agriculture, 2020a; Center for Plant Conservation, 2020). There are a limited number of published studies that describe the turkeybeard's habitat preferences. And there is only one published study that offered a predictive habitat suitability model for turkeybeard, which was situated in western and eastern Virginia and described a set of environmental variables that might correlate well with turkeybeard occurrences. These are elevation, slope, forest type, and fire frequency (Bourg et al., 2005).

To manage natural areas in a sustainable manner, habitat modeling approaches are necessary for evaluating current and desired future conditions of forests. Ideal management plans for public lands would have a conservation perspective, provide a set of sustainable goals, and employ habitat modeling approaches (Pecchi et al., 2019a; Pecchi et al., 2019b). For instance, forest management goals might state a need to increase plant species diversity and restore the natural integrity of a forested landscape. Therefore, a critical aspect of forest management is to use habitat models to gauge the development direction of a forest when management activities are employed (Edenius and Mikusiński, 2006).

Managing forest areas using sustainable approaches has become a worldwide issue (Liu et al., 2011). Over the past 30 years, scientists and experts have used habitat modeling approaches to make informed decisions concerning the management of forestlands for biodiversity conservation (Edenius and Mikusiński, 2006; Pecchi et al., 2019a) The U.S. national forests are currently managed under an approach that focuses on the sustainability of ecosystems (United States Department of Agriculture, 2020b). Therefore, habitat suitability models may be of value in assessing current and future trajectories of forest conditions. The U.S. has more than 150 national forests, and all their management plans are, in theory, designed to lead to sustainable systems. Some of these plans include habitat suitability modeling approaches.

Understanding the future dynamics and spatial distribution of forest ecosystems is a crucial issue for those concerned with the conservation of biodiversity (Rehfeldt et al., 2014). Forest ecosystems provide a wide range of ecosystem services to society. Thus, one goal of many public land natural resource programs is to develop sound management practices that indicate ecosystem sustainability will be achieved (Pecchi et al., 2019a). For example, when considering early seral stage regeneration activities, the use of adapted forest tree species (genotypes) and wise provenance selection (genotyping) has the potential to improve the resilience of forest systems (Ferrarini et al., 2016). Since the emergence of modern ecosystem modeling techniques in the last three decades, spatial data that include aerial and satellite imagery, vector feature databases, and national forest inventories are all considered vital resources for these efforts (Di Biase et al., 2018).

Before the mid-1980s, habitat modeling studies were limited due to the lack of comprehensive spatial datasets for climate and other factors useful for understanding and predicting forest conditions. Modern habitat modeling studies expanded in the late 1990s and early 2000s with the release of global climatic databases such as WorldClim, which provided the climatic data used in many habitat modeling studies to date (Hijmans and Graham, 2006). In addition, advances made in computer technology (hardware) over the last three decades, along with simultaneous advances in statistical algorithms (and associated computer programs), helped to make habitat modeling approaches easier to use, and thus, more widely applied in ecology and management. The species distribution modeling (SDM) approach was initially based on the ecological niche concept provided by Hutchinson around the 1950s and then refined by Booth et al. (1988).

Species distribution models are used to predict the most likely geographic distribution of a target species utilizing environmental data, and can also be used to forecast potential changes in distribution using Global Climate Model (GCM) outputs (Booth, 2018). These models can have many different names, including habitat models, niche-based models, habitat suitability models, environmental niche models (ENM), and ecological niche models (Pecchi et al., 2019b). Giles (1978) defines a Habitat Suitability Index (HSI) as an index of the carrying capacity of an area for a given species. The United States Fish and Wildlife Service (1981) quantified HSI as an organism's life requisites, using the structure, composition, and spatial components of habitat. The United States Fish and Wildlife Service HSI models were developed from 1980-1987, with the production of about 100 species-specific models. In the late 1980s, the United States and Wildlife Service emphasized the publication of models that were validated. Habitat suitability models (HSM) are observational strategies that relate field perceptions or museum-type species information about a species to natural indicator factors, based on a combination of measurably or hypothetically determined knowledge that best reflects the biological prerequisites of the species (Guisan et al., 2017). HSMs have found widespread application in species management, environmental impact assessment, abundance distribution, and ecological restoration (Brambilla et al., 2009; Imam et al., 2009). HSMs were first created as part of habitat evaluation methods to evaluate the effects of land management options on wildlife habitat (United States Fish and Wildlife Service, 1981).

It has been long accepted that habitat selection, habitat segregation, and limiting environmental factors are all interrelated concepts supporting species-habitat relationship modeling. However, researchers have been voicing concern over contradictory validation methods in studies (Roloff and Kernohan, 1999). Various explanations for these inconsistencies have been offered by scientists. Cook and Irwin (1985) suggest the problem relates to inadequate population sampling, Clark and Lewis (1983) suggest the problem arises from sampling in a limited range of habitat conditions, Bart et al. (1984) suggest modeled equations may not be not representative of the actual wildlife-habitat relationships, Cook and Irwin (1985) further suggest misinterpretation of results, Wiens (1986) suggests that the application of the model may be inappropriate, and Bender et al. (1996) suggest the problem may arise from inadequate consideration of data variability.

Regardless of these problems, habitat models allow quantification of habitat suitability as a proxy for the response of a species to an environment (Kliskey et al., 1999). In the United States, habitat suitability models became popular among species-orientated approaches to management and conservation planning in the 1980s, and a considerable number of models were developed for many species (Verner et al., 1986). Models such as these continue to be developed, and the development of habitat suitability indices has further been driven by advances in data processing capacity, analysis techniques and shifts in scientific paradigms, and increased attention to environmental heterogeneity and scale issues (Wiens, 2002). Models of habitat suitability use a variety of data types to describe habitat conditions of importance to different plant and animal species (Lele et al., 2013). Presence/absence models estimate the probability of species present within a given period at specific locations, using the information on both habitats use (species presence) and non-use (absence). Models of habitat suitability are derived from the theory of resource selection and are designed to estimate habitat selection by a species (Phillips et al., 2006; McDonald, 2013). Models for presence-only habitat suitability estimate the geographic distribution of potentially suitable sites from a representative sample of habitat measurements at sites that are actually observed to be used by a plant or animal species (Rotenberry et al., 2006).

There are associated indices of habitat suitability that can be developed by different types of models. Outputs from habitat models essentially represent habitat preference as described by a specific set of variables (Lele et al., 2013). Habitat suitability models usually seek to infer distributions of organisms in relation to management information and environmental relationships. Therefore, if different types of habitat models are applied to the same environmental range for data sampling of the same population, the outputs might be different but should be associated. Nonetheless, a lack of species occurrence data as an input to these models restricts a model's ability to provide accurate predictions (Thuiller et al., 2004; Morrison 2012). Furthermore, since habitat models quantify environmental relations differently, their outcomes may differ when used in different locations (Heikkinen et al., 2012; Latif et al., 2013).

Nonetheless, it is invaluable to evaluate habitat models with multiple criteria and independent datasets to verify their usefulness for guidance management. Independent data analysis is standard practice for assessing a model's predictive ability (Latif et al., 2015). Most often, for subsequent evaluation, a randomly selected subset of data is held back during the model development process (Guisan and Zimmermann, 2000). This subset is then used to validate model outcomes. However, models can also be tested against independently collected data from beyond the geographic scope of the original data used in model development to fully evaluate their predictive ability, or transferability (Heikkinen et al., 2012; Wenger and Olden, 2012; Bahn and McGill, 2013).

Given the ease with which habitat-related information can be derived from forest inventories, it is tempting to produce habitat suitability models that are a function of forest character and forest management history. This is also understandable when one considers the pace of habitat destruction and alteration in some areas of the world. Thus, to support sound and cost-efficient management, it is vital to develop, use, and evaluate habitat models as planning tools. The successful application of habitat models, therefore, involves consideration of different conceptual and methodological issues (Edenius and Mikusiński, 2006).

There are several different methods for developing HSMs, such as generalized linear model (GLM) (McCullagh and Nelder, 1989), classification and regression trees (CART) (Breiman et al., 1984), random forest (RF) (Breiman, 2001), artificial neural networks (ANN) (Rosenblatt, 1958), generalized additive model (GAM) (Hastie and Tibshirani, 1990) and maximum entropy (MaxEnt) (Phillips et al., 2006) methods. For many years, researchers have compared these models without reaching a consensus on which performs best under different conditions (Thuiller and Münkemüller, 2010). Therefore, it is often the decision of the analyst or modeler as to the method employed for generating a habitat model. The habitat suitability modeling process ideally follows five steps (Guisan et al., 2017):

- Conceptualization
- Data preparation
- Model calibration (fitting)
- Model evaluation
- Spatial predictions

A habitat suitability model for identifying areas with a high likelihood of supporting the eastern turkeybeard is important to the Talladega National Forest. Despite the lack of a protection plan (United States Forest Service, 2004), they are planning to collect more population data and more information about turkeybeard species for their future actions. Based on the limited knowledge about turkeybeard species in this research I used the only two available research, Bourg et al. (2005) and Bourg et. al. (2015), as a guideline. They found the most important variables for turkeybeard species are elevation, slope, fire frequency, and forest type. They also found turkeybeard is fire-adapted forest understory species. The conceptualization of a habitat suitability model is partly achieved in this section of this master's thesis using this limited knowledge. In the Methods section, the processes used to prepare the data, to develop, calibrate, and evaluate a habitat model, and to make spatial predictions. This will further contribute to the conceptualization of the model. I hypothesize that through the development of a habitat model, the highest valued habitat can be identified, and the correlation between the presence of the plant and the highly valued habitat is positive and relatively good.

CHAPTER 2

METHODS

This research entails developing a habitat suitability model for *Xerophyllum asphodeloides* (turkeybeard) for the Talladega National Forest in Alabama. Actual physical locations of *X. asphodeloides* were overlaid with several raster GIS databases that describe the physical environment to model the species-environment relationships and map potentially suitable habitats throughout the TNF. A maximum entropy model (MaxEnt), a generalized linear model (GLM), a generalized additive model (GAM), and a random forest (RF) model were each employed to produce probability distribution raster maps that may act to predict where conditions are most suitable for *X. asphodeloides*. The best model from GAM and GLM methods was reported here. The average of ten runs of the MaxEnt and the RF model is also reported. An ensemble model was then developed to indicate most suitable habitat areas.

2.1 Study Area

The lands that today form the Talladega National Forest are located in northeastern Alabama. They were mainly purchased as forest reserves from private landowners in the early 1900s through provisions contained in the Weeks Law of 1911 (P.L. 61-435, Chapter 186) to help regulate the flow of navigable streams in the headwaters of watersheds located mainly in the eastern states. Since its proclamation as a national forest in 1936, the Talladega National Forest has been managed by the US Forest Service (United States Department of Agriculture, 2012). The National Forest is physically separated into two areas, and these are further allocated to three

Ranger Districts: the Oakmulgee Ranger District, the Shoal Creek Ranger District, and the Talladega Ranger District. In all, the National Forest is located in 11 Alabama counties. However, the study area involves the Shoal Creek and Talladega Ranger Districts, which are located in Calhoun, Clay, Calhoun, Cherokee, and Talladega counties in Alabama. The lands of the Talladega National Forest are situated in the Talladega Mountains, and they are technically a part of Piedmont, and Ridge and Valley ecoregions of the United States (University of Alabama Department of Geography, 2020a) (Figure 1).

The climate of the Talladega National Forest is considered humid subtropical, with hot and humid summers and mild winters. In this area, the annual precipitation is about 1,260 mm, the annual maximum temperature is about 23.83 ℃, and the annual minimum temperature is about 10.16 ℃ (University of Alabama Department of Geography, 2020b). The national forest is composed of 95,276 ha of land and there are three main vegetation types: mixed coniferous and deciduous forests, deciduous forests, and coniferous forests. These vegetation types generally include broadleaf hardwood (deciduous), natural pine, and planted pine forests. Nearly 52% of the national forest is composed of forests dominated by four coniferous tree species: loblolly pine (*Pinus taeda*), longleaf pine (*P. palustris* Mill.), shortleaf pine (*P. echinata* Mill.), and Virginia pine (*P. virginiana* Mill.) (Stober et al., 2020).

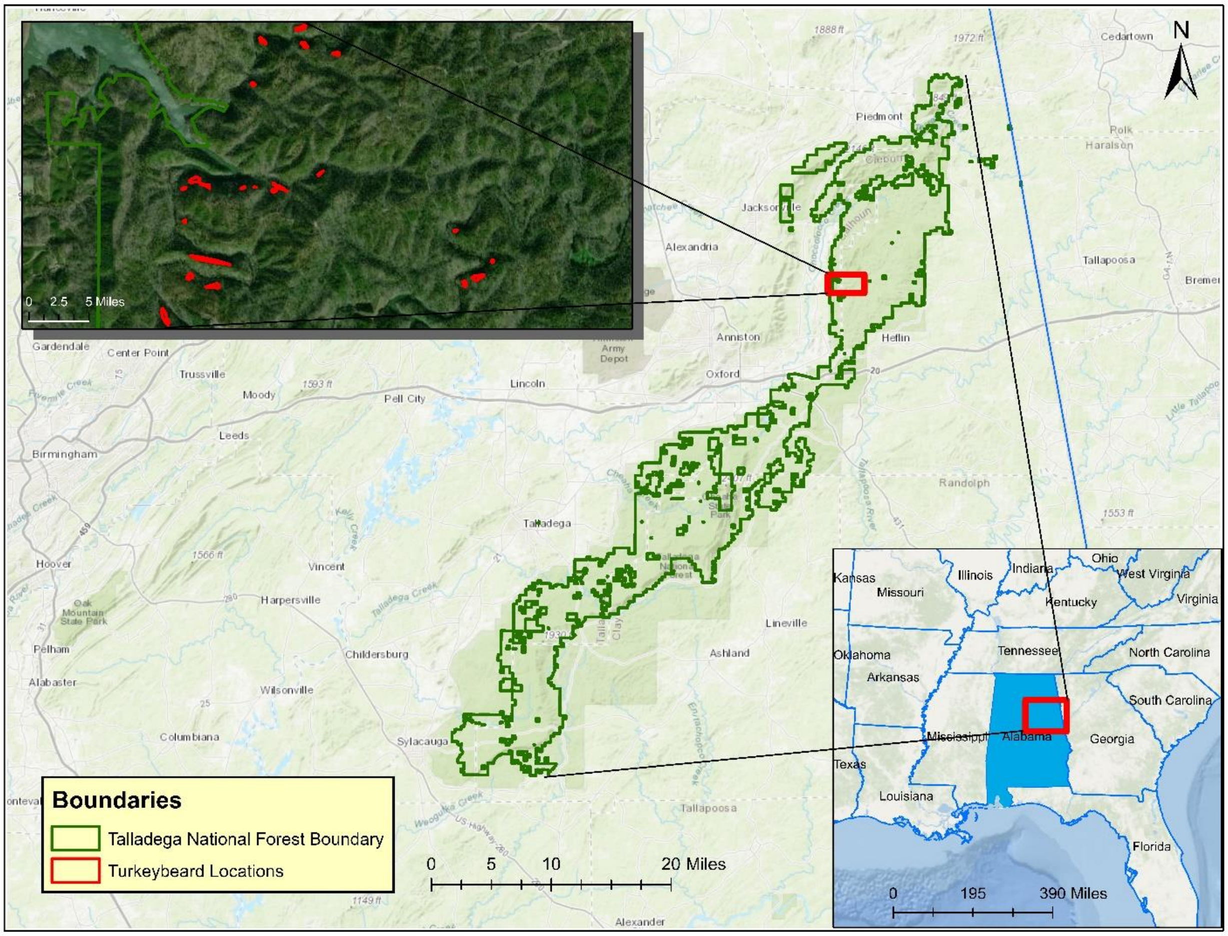

Figure 1.Talladega National Forest and turkeybeard occurrence locations in Talladega National Forest. (Basemaps sources: ESRI)

2.2 Occurrence Data

Xerophyllum asphodeloides (turkeybeard) is a rare and threatened plant species native to the United States and is mainly located in 10 eastern states (Figure 2). Several databases were examined to determine whether physical occurrence locations of the plant existed; however, there seems to be no database containing the plant's occupancy on the Talladega National Forest other than records from the Talladega National Forest. We obtained 22 occurrence locations of the species from the Talladega National Forest, consisting of polygons that represent patches of turkeybeard (Figure 1). The number of individual plants in these sites ranged from 28 to 1618, but individual plant locations were not available. These plant populations are 2.62 hectares in total in the study area. These occurrence polygons ranged between 18.94 and 5,022.23 square meters. These observations (patches of *Xerophyllum asphodeloides*) were collected by the national forest staff between 2015 and 2021.

In addition, we went to the field to collect occurrences points data from new field surveys and to check the already-identified occurrence places using a Trimble Juno T41 GNSS (Global Navigation Satellite System) receiver which has a horizontal accuracy of about 2 m in forested conditions (Lee et al., 2020). An ad-hoc survey was conducted for 7 hours on 1 day. With the assistance of a forest service employee, we visited 6 potential sites to determine the presence of turkeybeard. Three of these sites were already known turkeybeard populations. These sites were identified using a preliminary analysis conducted by another researcher (results unpublished). However, we were not able to find new occurrence data. We used 22 locations of the presence only of *Xerophyllum asphodeloides* from the national forest's observations. We confirmed the 3 already known occurrence location data using Google Earth Pro. In Google Earth Pro we checked our collected points whether they are matching with the occurrence's points collected from National Forest Service. The data provided by the national forest was in polygon format and using ArcMap 10.8 we extracted the occurrence as a point that represented the centroid of each polygon.

Figure 2. *Xerophyllum asphodeloides* (turkeybeard) in Talladega National Forest (Taken on 12.08.2020)

2.3 Environmental Variables

In this study, biotic and abiotic variables were used to develop a relationship with *Xerophyllum asphodeloides* occurrences. Twenty-eight abiotic and biotic variables were selected for this analysis: 3 of them are elevation, aspect, ground slope, 6 of them are soil-related (water content, soil organic material, soil percentage of clay, soil percentage of sand, soil percentage of silt, and soil PH), and 19 were bioclimate variables. The geographic information system (GIS) databases of 19 bioclimate variables were obtained from the WorldClim (WorldClim version 2.1) website (Fick and Hijmans, 2017, https://www.worldclim.org/data/worldclim21.html). A detailed explanation of these variables can be found in Appendix A. A 30 m digital elevation data was obtained from the United States Department of Agriculture Natural Resources

Conservation Service Geospatial Data Gateway (United States Department of Agriculture, 2014, https://datagateway.nrcs.usda.gov/GDGOrder.aspx), and this was used to calculate aspect and slope using the surface analysis tools in ArcGIS. Soil data were obtained from the USDA SSURGO (Soil Survey Geographic Database) database (United States Department of Agriculture, 2019). All of these data were resampled using ArcGIS to a 30 m spatial resolution (if not already available in that resolution) using the nearest neighbor technique which is a conservative resampling approach because it is good at preserving original pixel values. These variables are also projected to NAD 1983 UTM Zone 16N using the conversion tools within ArcGIS. The Talladega National Forest property boundary was used as an extent for all of these variables for analyses.

2.4 Habitat Suitability Modeling

For the modeling of suitable habitat for *Xerophyllum asphodeloides* in our study area, we employed 4 commonly used modeling methods; Generalized Linear Model (GLM), Generalized Additive Model (GAM), Maximum Entropy (MaxEnt), and Random Forest (RF). These models were used in this study to evaluate each variable's habitat suitability contributions to a *Xerophyllum asphodeloides* habitat model. After these predictions, these 4 different model predictions were exported from R Studio, and a final ensemble model was produced in ArcGIS. All 4 modeling efforts were conducted using R 3.5.2 software.

Variable Preparation and Selection

For the modeling processes, the extent, resolution, and projection of the raster databases were checked, and all NA (Not Available) values were omitted on these variables in R software.

For reducing multicollinearity problems, one of two strongly correlated variables was removed from the analysis by evaluating the correlation coefficient and the variance inflation factor (VIF). The Pearson Correlation Coefficient (Pearson's *r*) is a linear correlation coefficient that returns a value between -1 and +1. Based on the literature there is no certain rate of the threshold for a decision to remove one variable from the analysis that arose from a pair of highly correlated variables, as *r* = 0.8 is recommended by Menard (2002) and *r* = 0.7 is recommended by Green (1979). Further, after a correlation assessment, there may be still some missing correlated structure that cannot be seen in pairwise correlation (Guisan et al., 2017). To address these issues, the VIF test recommended by Harrell (2001), Guisan et al. (2006), Hair et al. (1998), and Guisan et al. (2017) was used. A VIF value below 10.0 suggested the model does not have a collinearity problem (Chatterjee and Hadi, 2006; Guisan et al., 2017), and a correlation below *r* = 0.7 suggested no strong correlation among the two variables. Once the highly correlated variables were removed from the analysis, each of the different modeling methods (described below) was applied separately. See Appendix B for the outcomes of this analysis.

Generalized Linear Model

GLMs are statistical extensions of linear models that do not impose unnatural scales of data and thus allow for structures of non-linearity and non-constant variation in the data. GLM is a generalization of multiple regression analysis that uses a binomial distribution and a logistic relation to suit polynomials of higher degrees of complexity greater than linear (Hirzel et al., 2001). In GLMs the mean of the response is determined by the linear predictor (McCullagh and Nelder, 1989). They are based on an implied relationship between the mean of the response variable and the explanatory variable's linear combination (Hastie and Tibshirani, 1990; Guisan et al., 2002). For the GLM analysis, 1000 random pseudo-absences were generated, 90% of these points were used as training points and 10% were used for validation testing. For the model selection and generation, the *stepAIC* function, stepwise regression analysis to select the best model based on their AIC score, was used in R and the best model was selected. This model was run 10 times and the solution with the lowest AIC score (the best) was the one selected.

Generalized Additive Model

Instead of being algorithm-driven, GAMs are data-driven; that is, the resulting fitted values do not originate from an *a priori* model (Yee and Mitchell, 1991). The only fundamental statement made is that the internal functions are additive, and the modules are smooth. GAMs employ "smoother" algorithms, which automatically match response curves "as closely as possible" to the data provided the degree of smoothing allowed (Guisan et al., 2017). GAMs are semi-parametric extensions of GLMs (Guisan et al., 2002). Thus, instead of being constrained by the shapes available in a parametric class, GAMs allow the data to decide the form of the response curves (Hastie and Tibshirani, 1990; Yee and Mitchell, 1991; Guisan et al., 2002). For the GAM analysis, 1000 random pseudo-absences were generated, and 90% were used as training points while 10% was used for validation testing. For the model selection and generation, the *stepGam* function was used in R, which is the stepwise regression analysis to model selection based on AIC scores. In a GAMs model, an ANOVA test is also applied to select the best solution. This model was run 10 times (like the GLM) and the solution that had the lowest AIC score (the best model) was selected.

MaxEnt

MaxEnt estimates a target probability distribution by finding the probability distribution of maximum entropy, which is subject to a series of restrictions that represent our imperfect knowledge about the target distribution (Phillips et al., 2006). MaxEnt calculates the likelihood that a species will be present based on reports of existence and produces background points randomly by estimating the maximum entropy distribution. GLMs, GAMs, and machine learning techniques such as MaxEnt have clear parallels with some existing methods for predicting species distributions (Phillips et al., 2006). Although MaxEnt is similar to GLMs and GAMSs, the GLMs and GAMs are discriminative methods, and they require both occurrence and absence data. However, MaxEnt is a generative method, it can operate without absence data (Phillips et al., 2006) and it may also be effective with a small amount of training data (Hernandez et al., 2006; Phillips et al., 2006; Phillips and Dudik, 2008; Benito et al., 2009). For the MaxEnt analysis, 1000 random background points were generated. Of these, 90% were used as training points and 10% were used for validation testing. MaxEnt was run 10 times like the GLM and GAM methods. But unlike the GLM and GAM models, after the 10 runs, the MaxEnt results were averaged and used for prediction and generating a suitability map.

Random Forest

RF is a form of machine learning algorithm which provides a metric of variable significance, both for classification and for regression (Garzón et al., 2006), and it executes a combination of tree predictors (Breiman, 2001). The RF is a non-parametric tree-based approach that involves a large number of classification and regression trees that are expanded regression tree models. Also, random forests can be created to search for overfitting, by using stochasticity in the tree-building process as well as at each node (Breiman, 2001). Three parameters must be taken into account when creating an RF: the proportion of sub-samples, the number of optimal estimators in each node, and the optimal number of trees (Shataee et al., 2012). For the RF analysis, 1000 random background points were generated, and 90% were used as training points, while 10% were used for validation testing. RF was run 10 times like the other methods. Like the MaxEnt model, the outcomes from 10 runs were averaged and used for prediction and generating a suitability map.

Final Habitat Suitability

To generate the final suitability modeling map ArcGIS was used. All of the modeling (MaxEnt, GLM, GAM, RF) predictions were exported from R Studio to ArcGIS to generate the final most probable suitable areas. All predictions from the 4 different models were classified as 0 (Unsuitable) and 1 (Suitable) in ArcGIS. To determine suitable and unsuitable habitats, each model was projected to the whole study area, creating a raster map of predicted suitability values. A threshold needs to be determined for the predicted values in order to classify those outputs as suitable and unsuitable. Thresholds were calculated based on the evaluation for each model, which used the sensitivity-specificity sum maximization approach recommended by Liu et. al. (2005). In this approach, the highest threshold is chosen, at which the sum of the sensitivity (true positive rate) and specificity (true negative rate) is highest. However, these approaches were not applied to final presence-absence maps and the combined final habitat suitability map. Instead, a threshold of 0.5 was applied, following the subjective-based approach, which is widely used in ecology and utilizes a fixed value of usually 0.5 (Manel et al., 1999; Bailey et al. 2002). These areas were classified as 1 (Suitable) to represent the most probable suitable areas and 0 (Unsuitable) to represent the most probable absence locations for the turkeybeard from each model. After this reclassification step, a raster calculator was used in ArcGIS to generate the final suitable maps. Two different approaches were used to generate final suitability maps. For the first approach, a conditional statement was applied to generate the final highest probable suitable and unsuitable locations in the raster calculator. The condition stated that the pixel value for each model had to be 1. These four statements inside of the condition were joined together by using the “and” operator. Thus, each model’s suitable habitat (value of 1) was retained in the output only if all models had a value of 1 for that pixel. This produced only values of 1 where all four models agreed that the pixel was suitable and produced values of 0 otherwise. Therefore, the first ensemble map was generated 0 (Unsuitable) and 1 (Suitable) sites. These final suitable sites with a value of 1 represented the most suitable areas for turkeybeard due to agreement from all four models. A second approach was applied using a raster calculator. Different from the first approach, all predictions were summed, which created 5 different levels with values 0, 1, 2, 3, 4. These numbers represent places not predicted by any model (0), places predicted by only one model (1), places predicted by 2 models in which both overlapped or agreed (2), places predicted by 3 models in which all 3 overlapped or agreed (3), and places predicted by all 4 models in which all 4 overlapped or agreed, respectively. The value of 4 represents the most highly suitable areas and the value of 0 represents the most unsuitable areas for the turkeybeard species.

2.5. Model Evaluation

To detect which environmental variables were the dominant factor for models, their variable contribution and coefficients were checked using models R packages; *randomForest, glm, gam* and *maxent*. An ANOVA test was employed in the GAM model to measure which variables seem to be the dominant ecological factors determining the potential distribution of the *Xerophyllum asphodeloides.* For the model evaluation, the *evaluate* function from the *dismo* package was used in R Studio for each model. The AUC, which is the area under the receiver operating characteristic (ROC curve), was used as the model's predictive performance metric (Wei et al., 2018). An AUC < 0.5 describes a model that performs worse than chance and this occurs rarely in reality since the 0.5 level refers to pure randomness. The performance of the model was rated as failing (0.5-0.6 AUC value), poor (0.6-0.7), fair (0.7-0.8), nice (0.8-0.9), or great (0.9-1.0) (Swets, 1988). The closer an AUC value to 1.0, the better the model efficiency.

After generating prediction maps for all four models in R, all maps were imported into ArcGIS and the final habitat suitability map was generated using ArcMap 10.8. A final visual check of the prediction was conducted in ArcMap. All areas predicted as suitable were compared with already known turkeybeard populations to assess overlap. Twenty-two occurrence locations (polygons) centroided and these points were used for the accuracy of all predictions. These centroids were checked to see whether actual locations overlapped with predictions as suitable habitat or not, and their overlapping percentage was calculated for each habitat suitability prediction.

CHAPTER 3

RESULTS

During this research, I was unable to collect new occurrence places for the *Xerophyllum asphodeloides*. I traveled to the study site in Fall 2020 to record the location of new occurrences and we visited several sites on the Talladega National Forest using an ad-hoc survey for 7 hours on 1 day, but we were unable to locate new places hosting the species. This was due to the rareness of the species and the wide range of the Talladega National Forest. I expected to have several visits to the Talladega National Forest, but due to the Covid-19 pandemic, access was limited and required an escort. For these reasons, I used only the Talladega National Forest's detected presence points (22 occurrences) for my research.

Sixteen environmental variables were eliminated based on Pearson's $r > 0.7$ and VIF > 10, leaving 12 variables to be used in the models (Table 1) (Figure 3).

Table 1. VIFs of the 12 variables that remain after eliminating some based on Pearson's *r* and VIF scores.

Variables	VIF Scores
Elevation	5.371157
Slope	1.215588
Aspect	1.029485
Soil available water content	1.972491
Soil organic material	2.019679
Soil percent clay	1.902937
Soil percent silt	2.991156
Soil percent pH	2.450709
Bio3 (Isothermality)	3.579550
Bio4 (Temperature seasonality)	4.743466
Bio8 (Mean temperature of wettest quarter)	1.605334
Bio15 (Precipitation seasonality)	2.740790

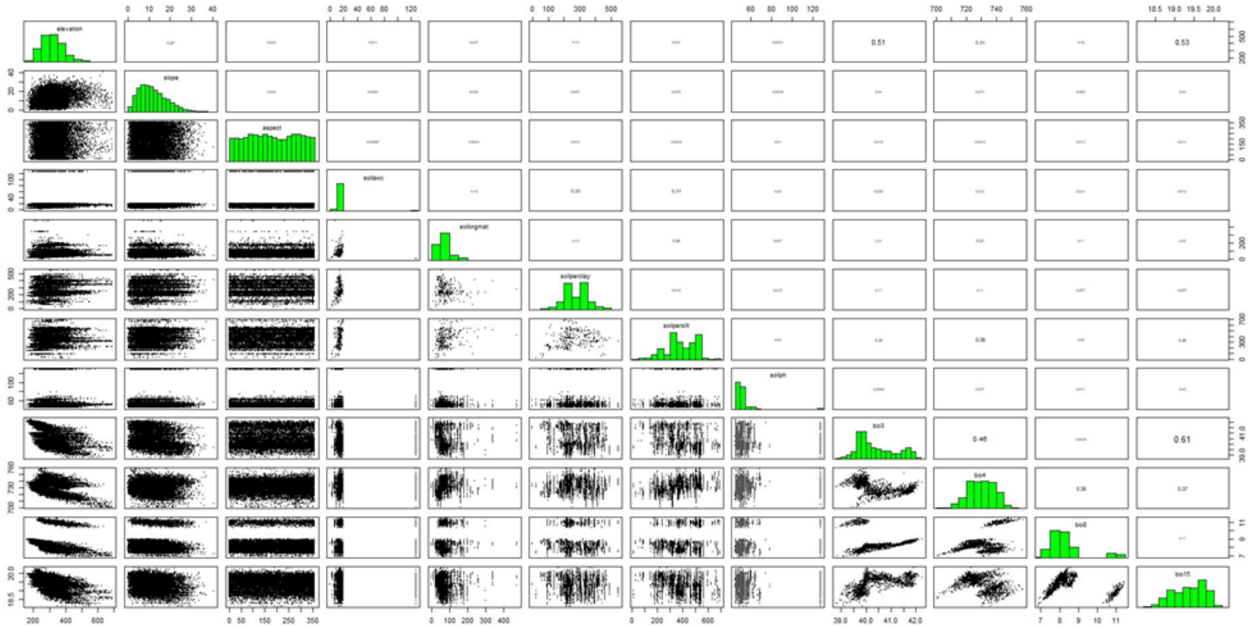

Figure 3. The correlation matrix for remaining variables whose correlations are under 0.7.

After 10 runs of MaxEnt using all 12 variables in each model, all runs were averaged, and they were evaluated to generate the prediction maps and AUC scores to evaluate the model's success. The average AUC score was 0.99 (Table 4). The ground slope had the highest contribution to the model (Figure 4). The second highest contributing variable was Bio4 (Temperature seasonality), and the third was Bio8 (Mean temperature of wettest quarter). The least contributing variables were aspect, soil available water content, and soil pH. Based on the model evaluation's success the MaxEnt two prediction maps were generated. One was based on the raw values (Figure 5) to show a range of relative habitat suitability, and the other was based on the suitable/unsuitable habitat map which was generated using a threshold value of 0.5 (Figure 6).

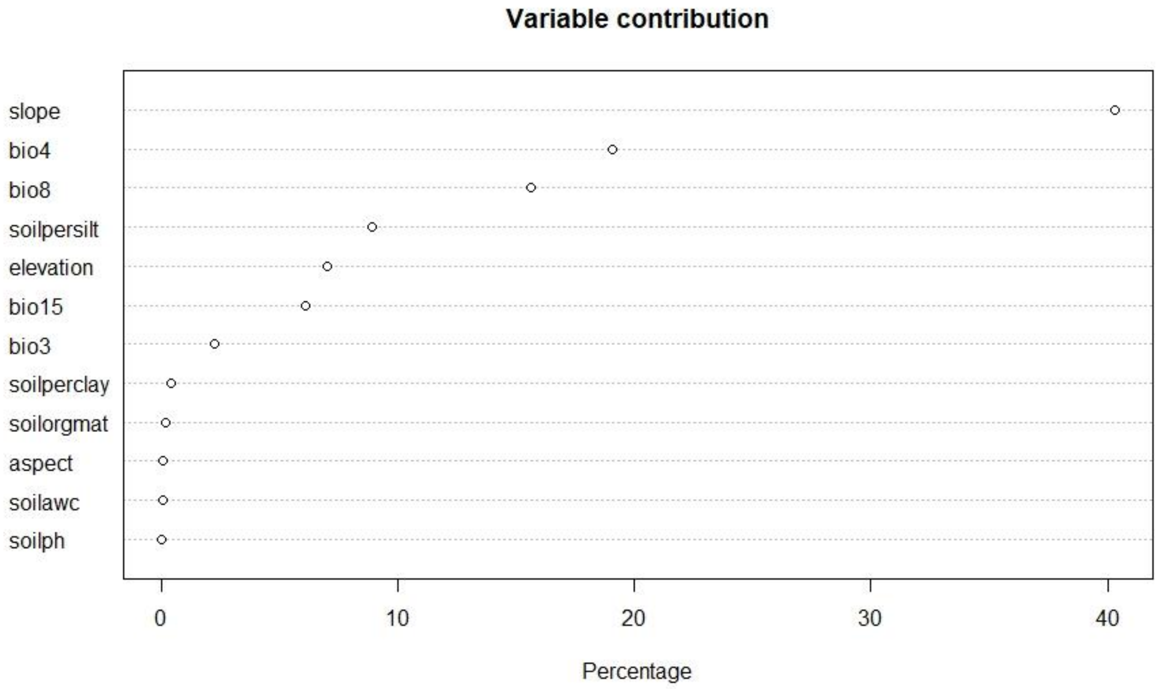

Figure 4. Contribution of the environmental variables in MaxEnt.

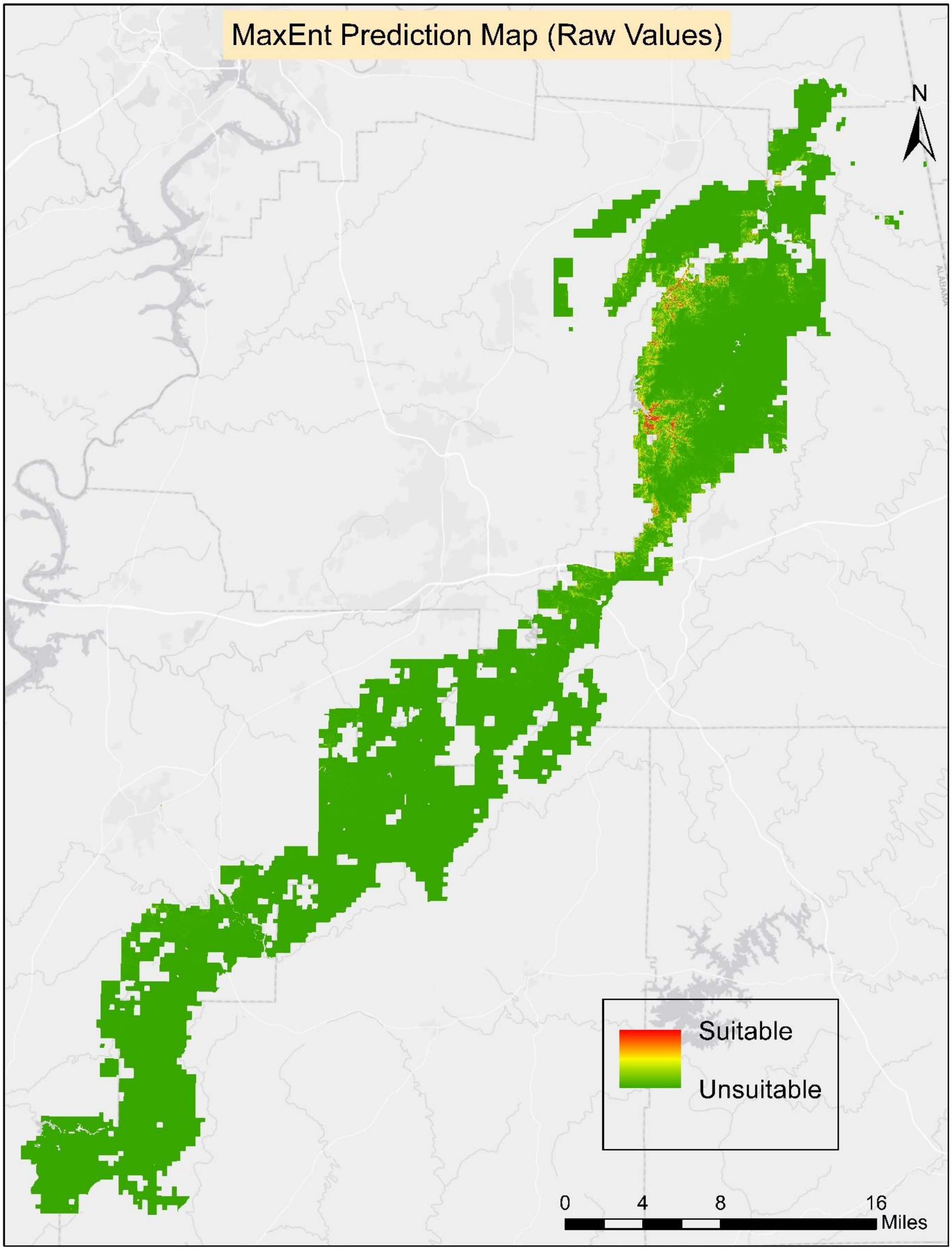

Figure 5. MaxEnt prediction map based on raw values.

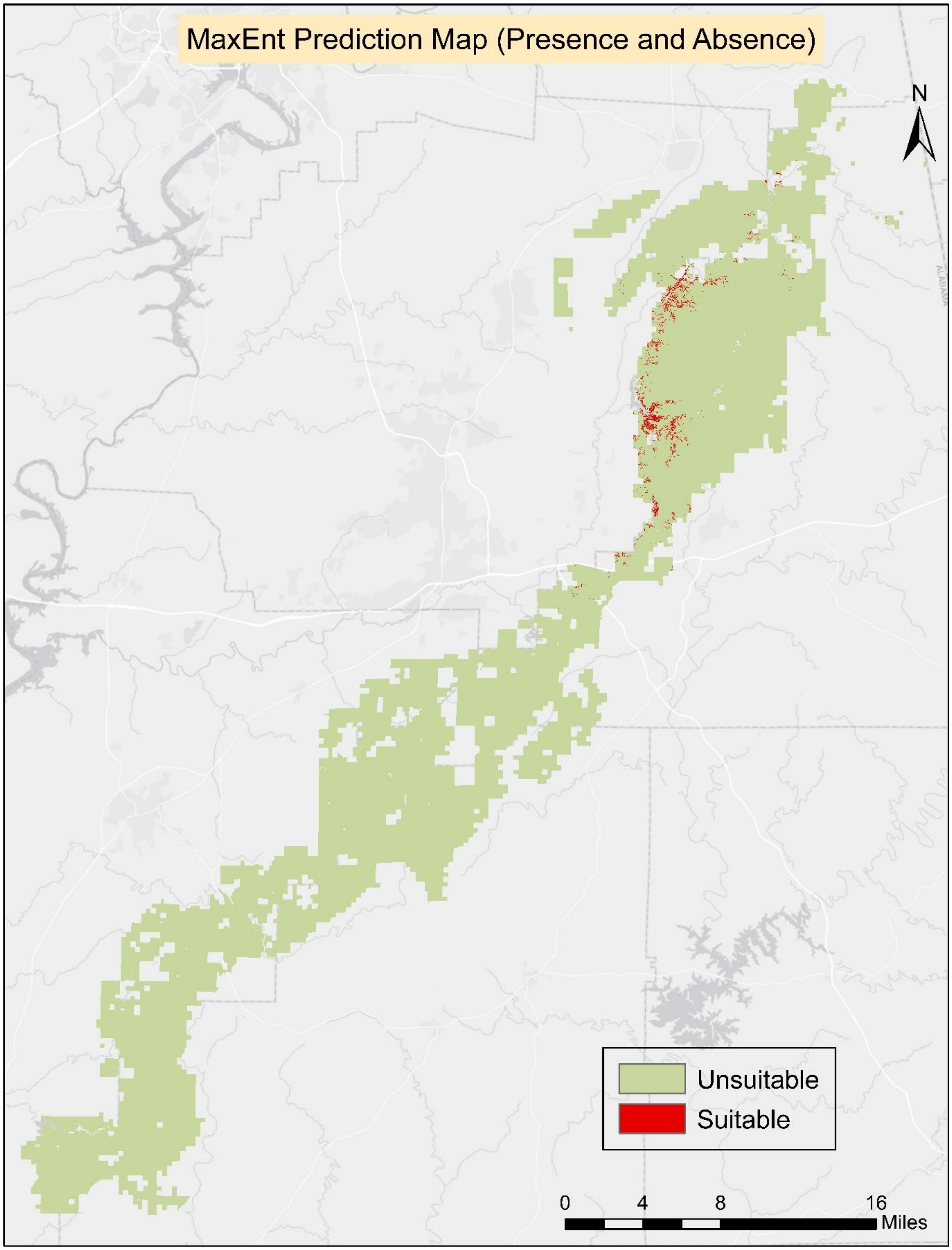

Figure 6. MaxEnt prediction map based on presence and absence.

For the GLM analysis, a stepwise logistic regression analysis was used to determine the best model based on AIC scores. To determine the best model, stepwise regression analysis ran 10 times, and the best AIC score was found at 78.82 after these 10 runs. The best GLM model included 7 environmental variables: ground slope, soil percent silt, Bio4 (Temperature seasonality), Bio8 (Mean temperature of wettest quarter), soil available water content, soil percent clay, and elevation. Based on the coefficients table for the variables on the best model, significant variables included slope, Bio4 (Temperature seasonality), soil percent clay, and soil available water content (Table 2). The remaining 3 variables (soil percent silt, Bio8 (Mean temperature of wettest quarter), and elevation) were found not to be significant. Based on the best model, the AUC score was 0.96 (Table 4). Two different prediction maps were generated. One was based on the raw values (Figure 7) and the other was based on presence and absence location maps of the species in the study area, which was generated using a threshold value of 0.5 (Figure 8).

Table 2. Coefficients Table for Generalized Linear Model.

Variables	Estimate	Std. Error	T Value	Pr (>|t|)
(Intercept)	-348.11543	103.78129	-3.354	0.000796*
Slope	0.25957	0.05415	4.794	0.000002*
Soil percent silt	0.02112	0.01406	1.503	0.132888
Bio4	0.47221	0.13442	3.513	0.000443*
Bio8	-1.26474	0.93523	-1.352	0.176271
Sawc$^{\alpha}$	-0.80335	0.36386	-2.208	0.027254*
Soil percent clay	0.03952	0.01441	2.742	0.006109*
Elevation	-0.02153	0.01303	-1.652	0.098472

* Based on the significance level of 0.05.

$^{\alpha}$ Sawc = Soil available water content.

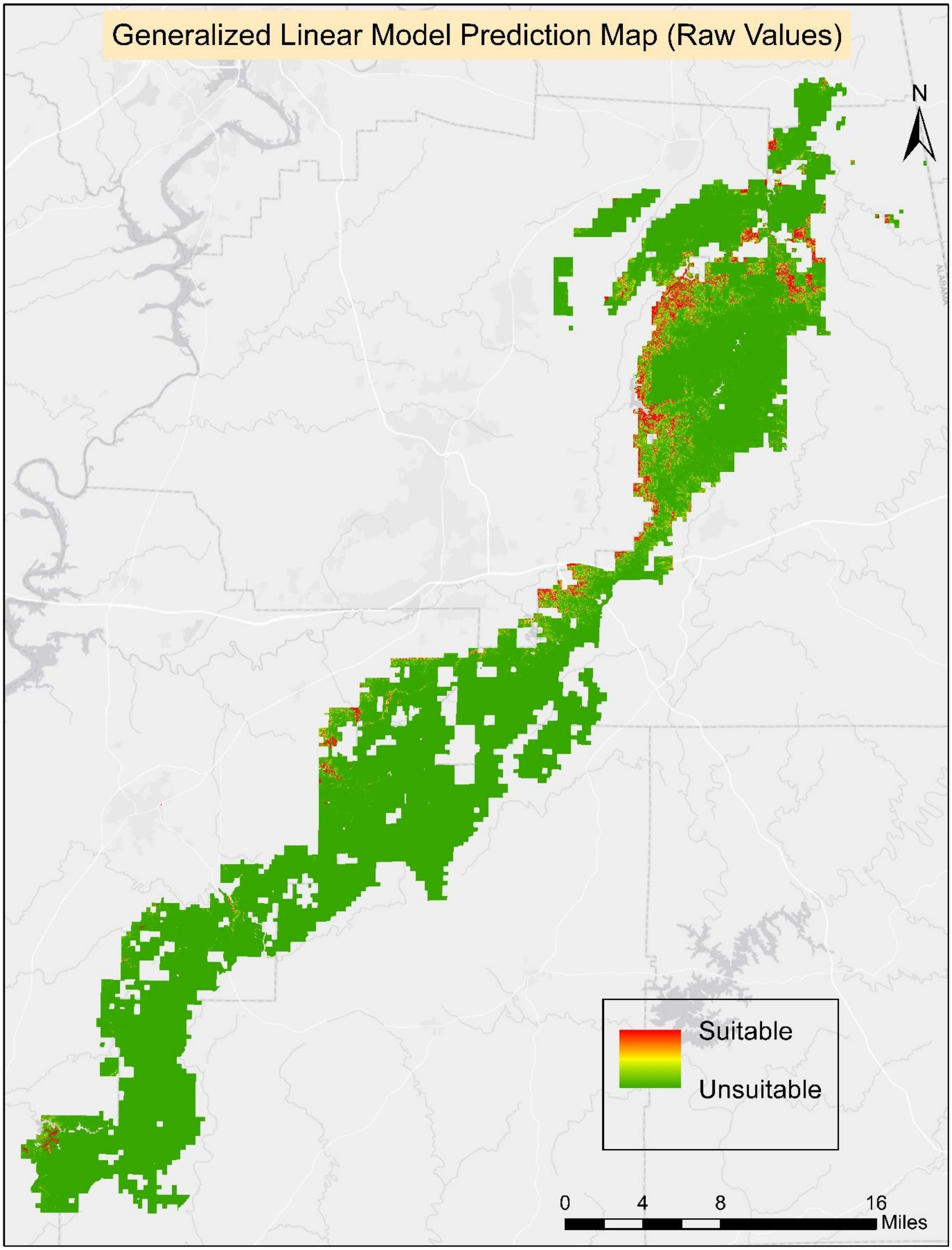

Figure 7. Generalized Linear Model prediction map based on raw values.

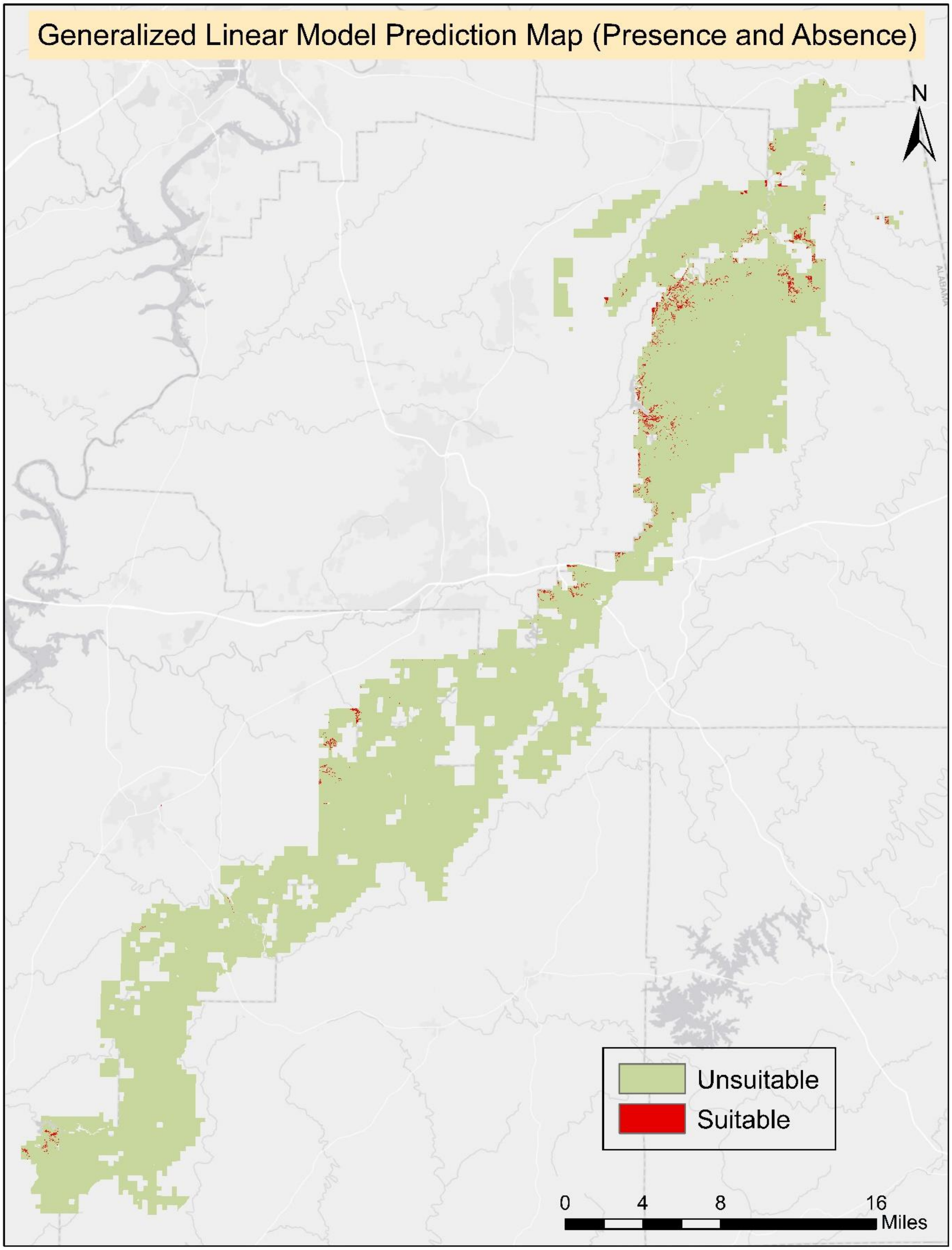

Figure 8. Generalized Linear Model prediction map based on presence and absence.

For the GAM analysis, a stepwise logistic regression analysis was used to determine the best model based on AIC scores. To determine the best model, stepwise regression analysis ran 10 times, and the best AIC score was found at 57.40 after these 10 runs. The best GAM model included 5 environmental variables: elevation, ground slope, soil percent clay, Bio4 (Temperature seasonality), and Bio8 (Mean temperature of wettest quarter). Based on the ANOVA parametric effects for the variables on the best model, all the variables' p-values were significant except the Bio4 (Temperature seasonality) (Table 3). Based on the ANOVA for nonparametric effects Bio4 (Temperature seasonality) and Bio8 (Mean temperature of wettest quarter) was found significant; however, elevation, ground slope, and soil percent clay were found insignificant for the nonparametric effects. The best GAM model AUC score was 0.98 (Table 4). It shows that the modeling process was successful. Two different GAM prediction maps were generated, one was based on the raw values (Figure 9) and the other was based on presence and absence locations maps of the species in the study area which was generated using a threshold value of 0.5 (Figure 10).

Table 3. Coefficients Table for Generalized Additive Model.

Variables	Sum Sq[a]	Mean Sq[b]	F Value	Pr (>F)
Elevation	4.872	4.8720	96.1326	< 2.2e-16*
Slope	12.019	12.0188	237.1506	< 2.2e-16*
Soil percent clay	2.312	2.3123	45.6260	2.547e-11*
Bio4	0.142	0.1419	2.7997	0.0946236
Bio8	0.685	0.6854	13.5239	0.0002493*

* Based on the significance level of 0.05.

[a] Sum of squares of variation.

[b] Mean of squares.

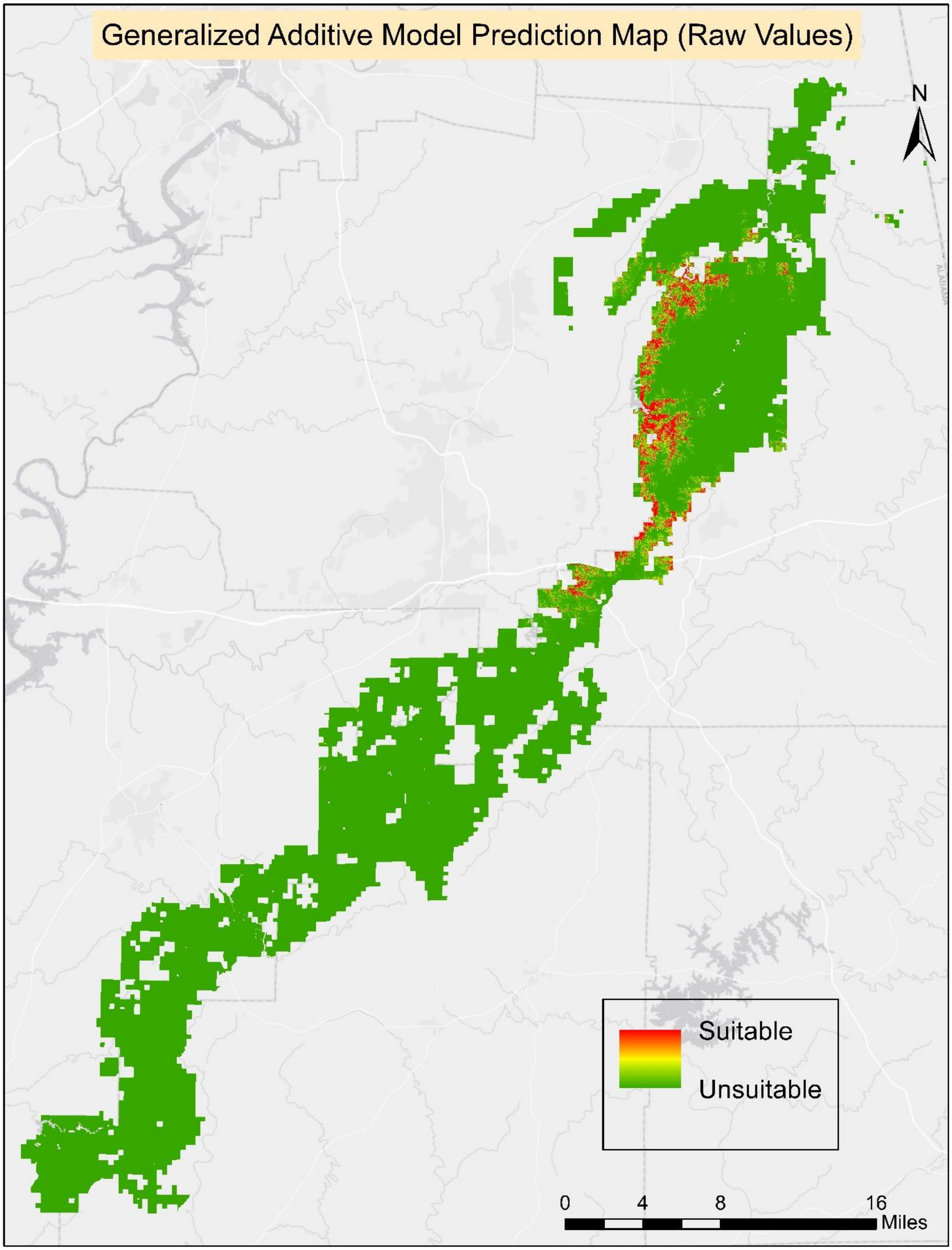

Figure 9. Generalized Additive Model prediction map based on raw values.

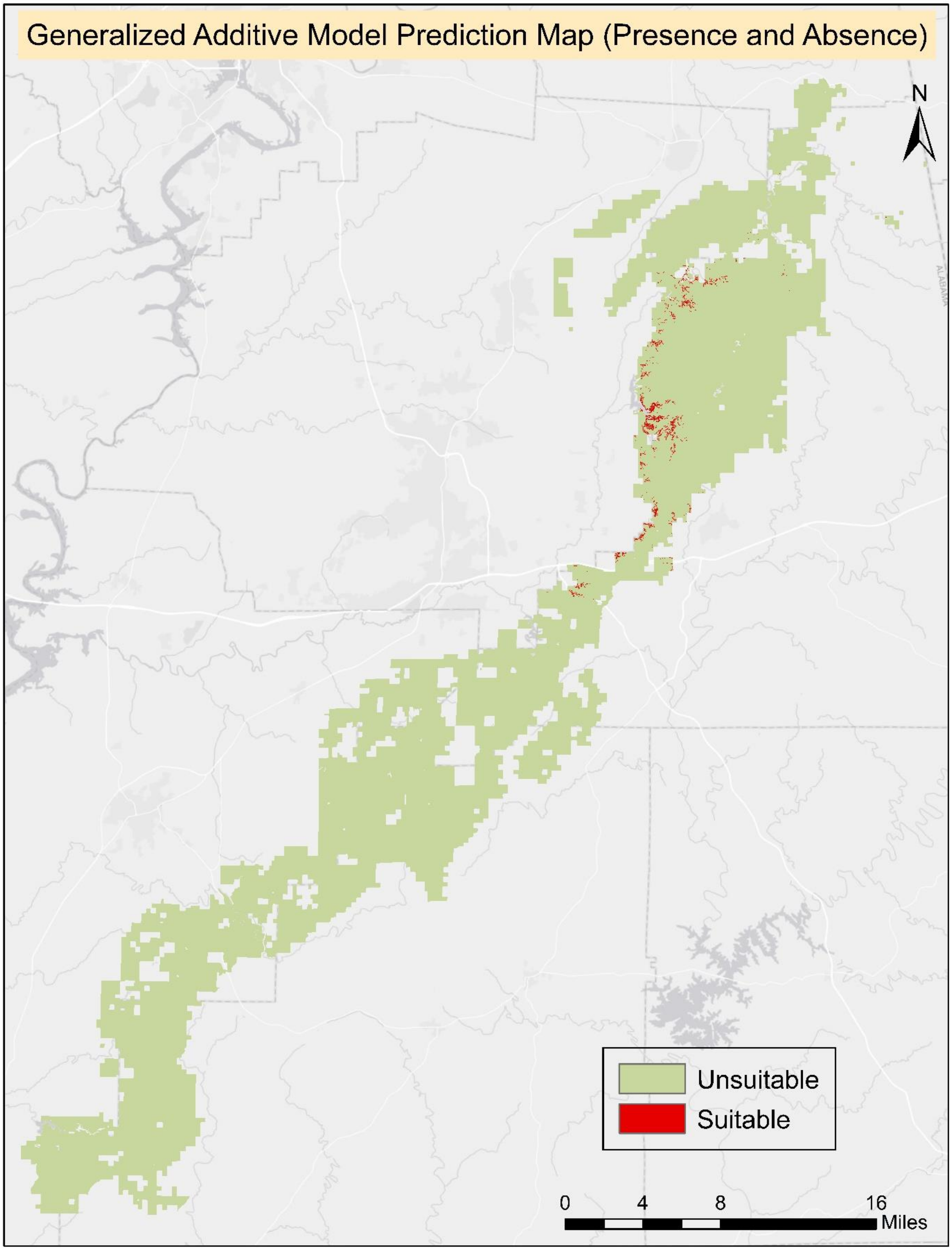

Figure 10. Generalized Additive Model prediction map based on presence and absence.

All 10 runs of RF were averaged, and they were evaluated to generate prediction maps and the AUC score to evaluate the model's success. The average AUC score was 0.99 (Table 4). All 12 variables which were selected after the correlation check was used in the RF modeling process. The ground slope was found as the most contributive variable in the model. The second most contributive environmental variable detected was elevation, and the third was Bio4 (Temperature seasonality). The least contributive variables detected were soil organic material, soil pH, and aspect. The detailed variable contribution given is in Figure 11. Again, two prediction maps were generated, one was based on the raw values (Figure 12) and the other was based on the presence and absence locations maps of the species in the study area which was generated using a threshold value of 0.5 (Figure 13).

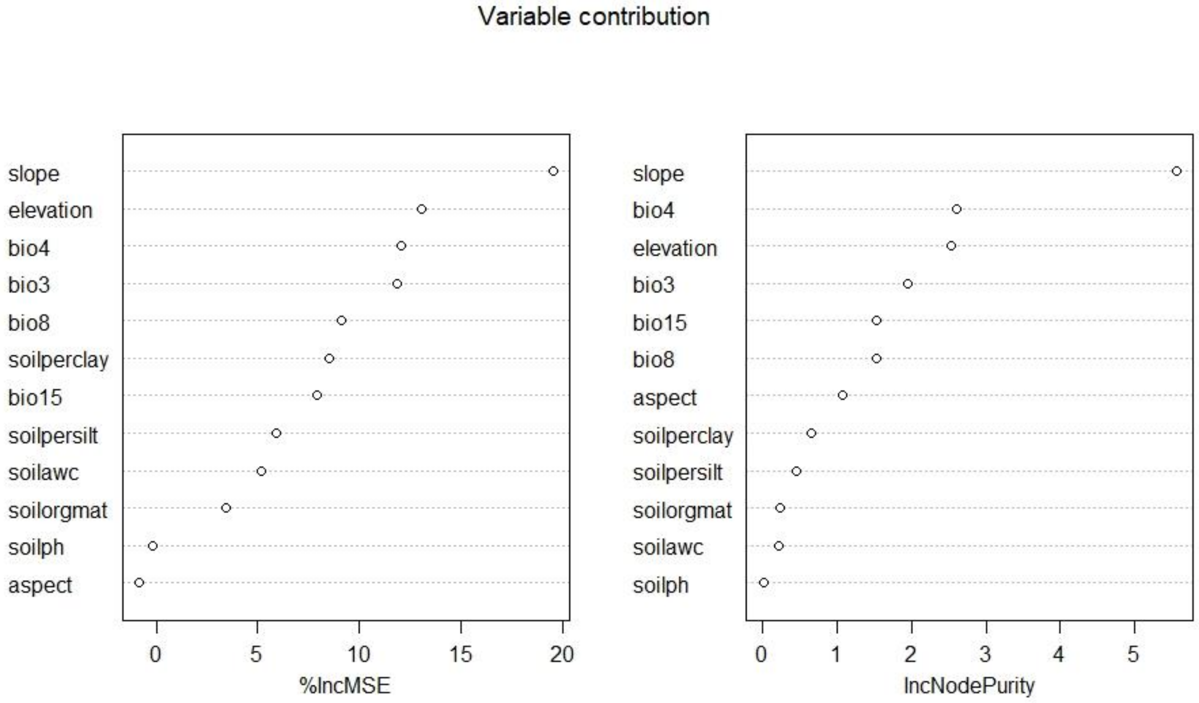

Figure 11. Contribution of the environmental variables in Random Forest.

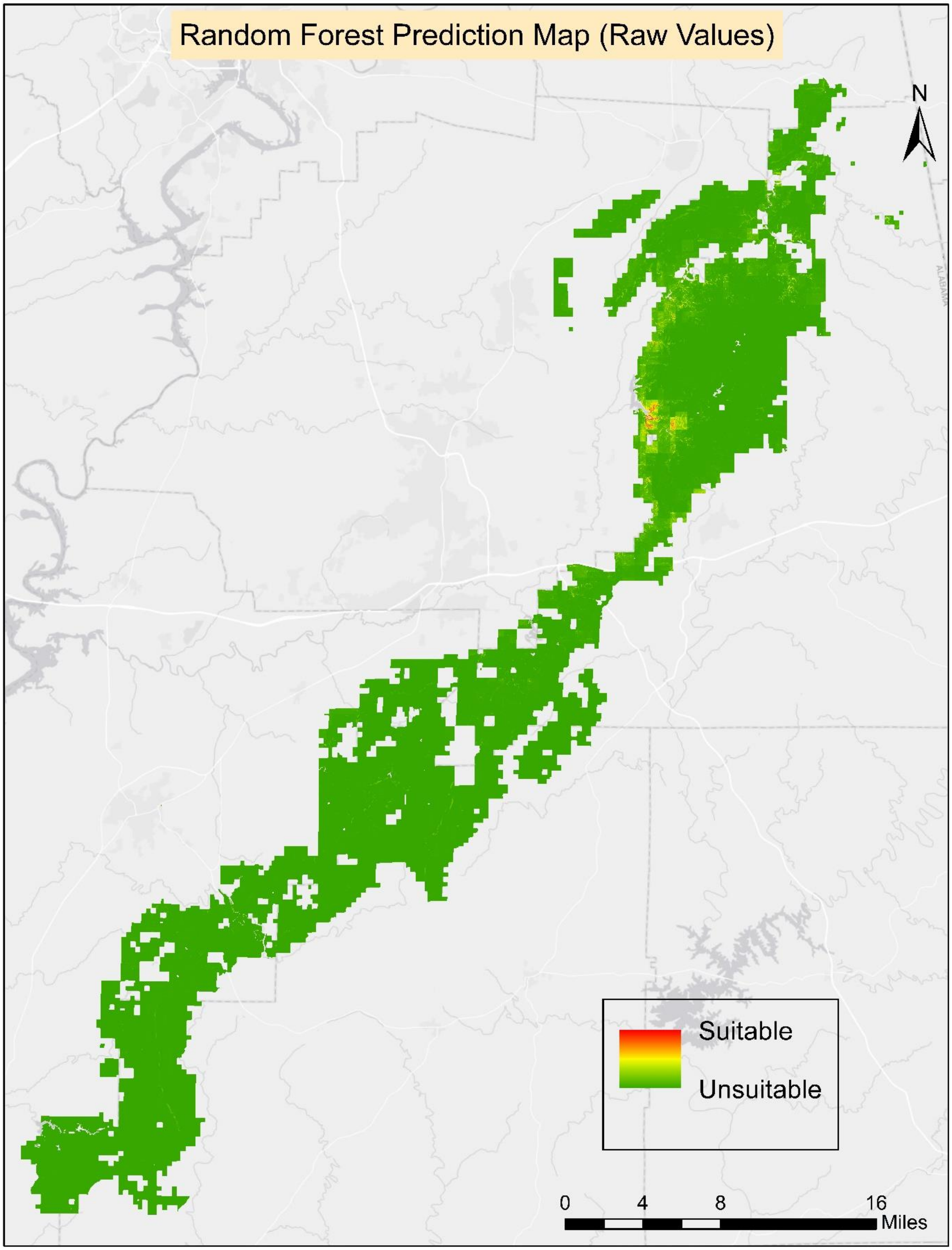

Figure 12. Random Forest prediction map based on raw values.

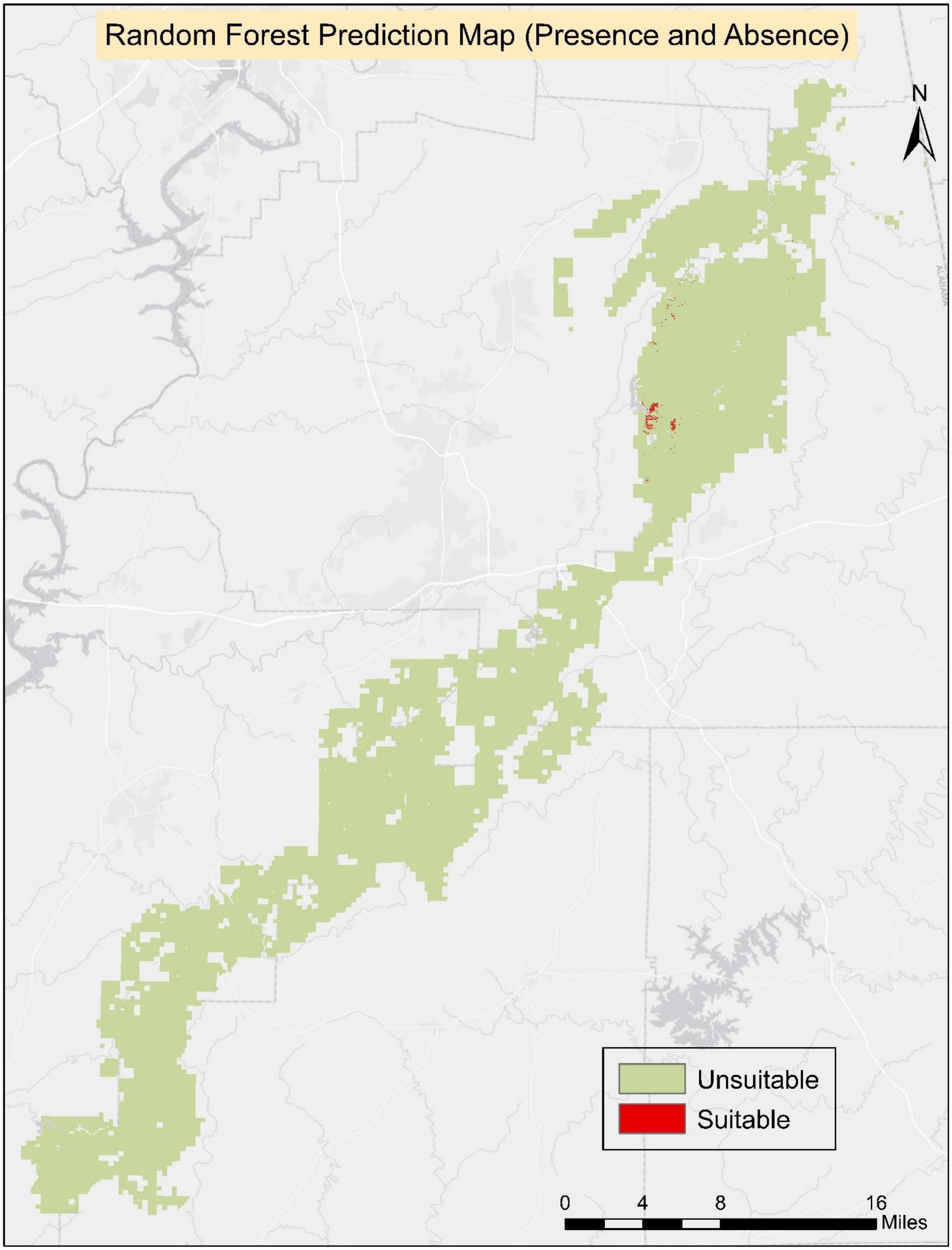

Figure 13. Random Forest prediction map based on presence and absence.

Table 4. AUC Scores of All Modelling Methods.

Models	AUC
Maxent	0.99
GLM	0.96
GAM	0.98
RF	0.99

Based on the evaluation of all modeling approaches, the highest AUC score was 0.99 for the MaxEnt and RF analyses. The lowest AUC score was 0.96 for the GLM (Table 4), suggesting that the models were generally successful in classifying presence and pseudo-absence data for *X. asphodeloides* on the Talladega National Forest with the given environmental variables.

Based on our modeling process 97.01% (92,428.2 ha) of the land was assigned a predicted value. Leaving only 2.99% (2,847.8 ha) of the land with no data values as a result of the input data. The suitable areas based on each model prediction were calculated and the highest suitable areas in the 4 models were found as 851.49 ha (0.92%) on GLM predictions. The lowest amount of suitable area was found as 114.93 ha (0.12%) of the national forest through RF (Table 5). The final habitat map was generated from the composite of 4 methods, which suggest 79.02 ha (0.09%) of suitable areas for turkeybeard and 92,349.18 ha (99.91%) is unsuitable (Figure 14). However, the other approach for generating the final habitat suitability map found that the unsuitable areas were over 1,000 ha less, at 91,127.3 ha (98.59%) (Figure 15, Figure 16). The prediction area from the agreement of 3 models was 164.61 ha (0.79%), the prediction area from the agreement of 2 models was 329.49 ha (0.36%), and the prediction area from a single model with no agreement from other models was 727.83 ha (0.18%) (Figure 15, Figure 16). The already known presence locations amount is 2.62 ha and therefore nearly 76 ha new potential occurrence locations are suggested.

Table 5. Suitable areas based on the models for the turkeybeard species in Talladega National Forest.

Model	Area (ha)	Percent (%) of the national forest
MaxEnt	708.84	0.77%
GLM	851.49	0.92%
GAM	521.46	0.56%
RF	114.93	0.12%

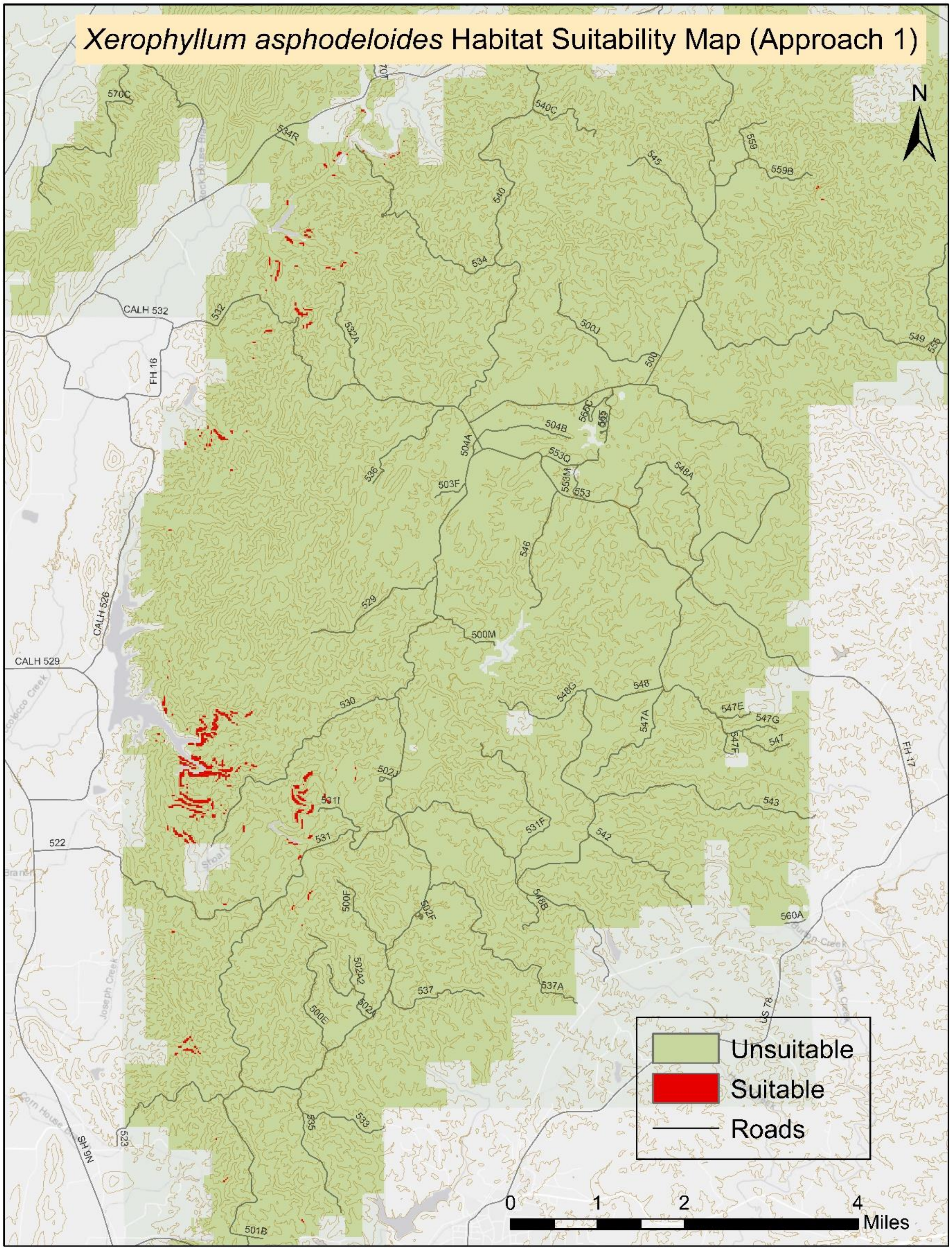

Figure 14. *Xerophyllum asphodeloides* (turkeybeard) habitat suitability map based on approach 1. (Basemap source: ESRI)

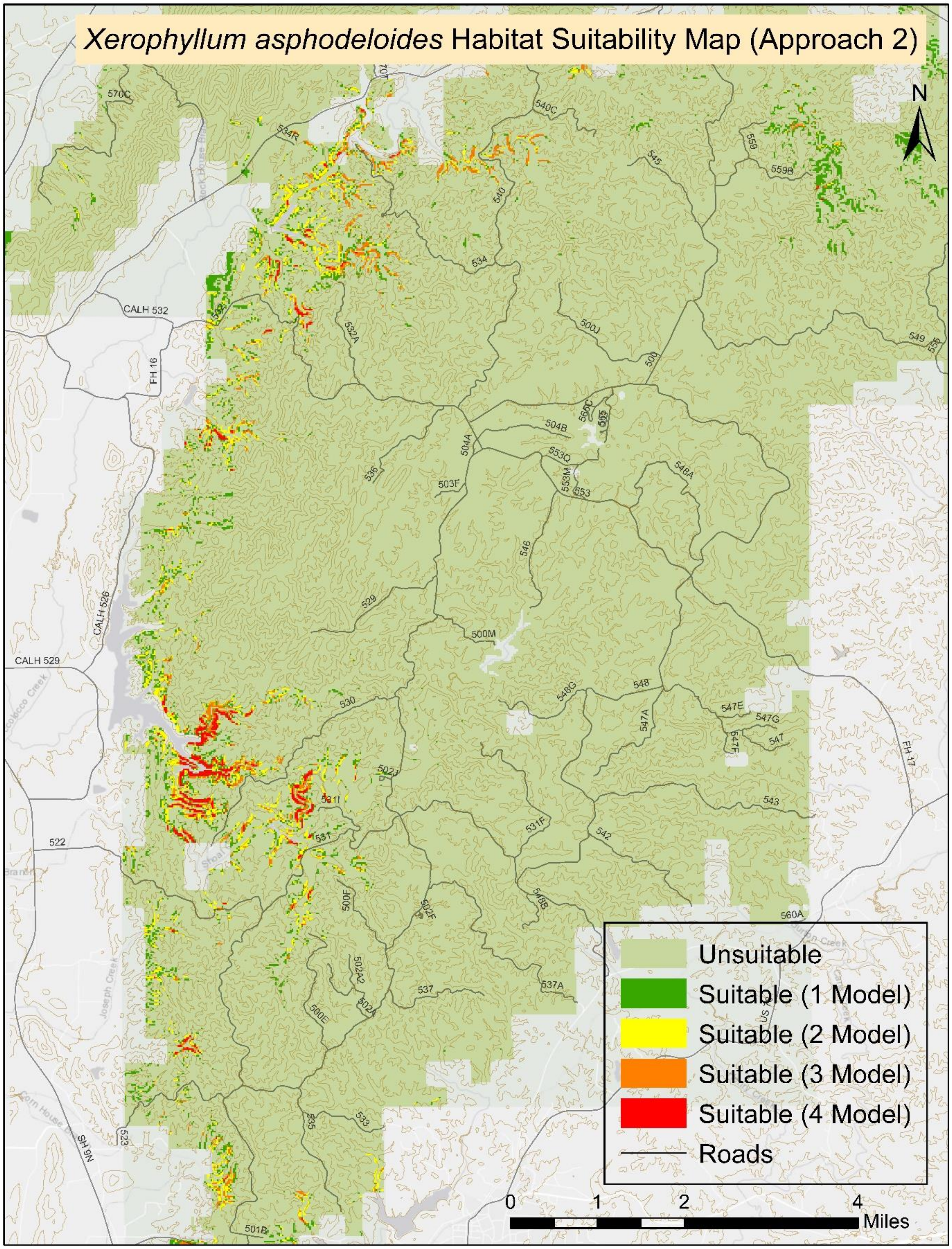

Figure 15. *Xerophyllum asphodeloides* (turkeybeard) habitat suitability map based on approach 2. (Basemap source: ESRI)

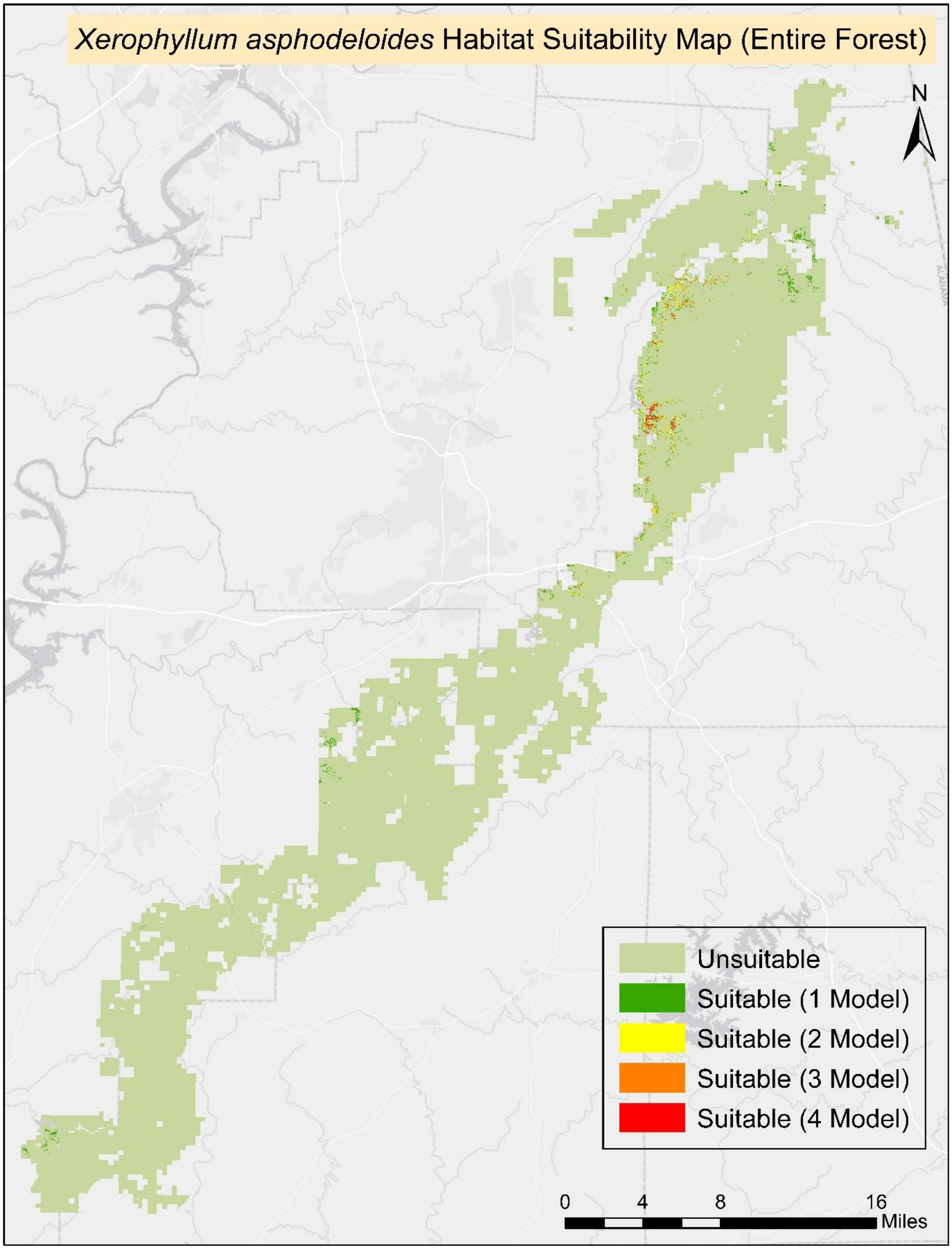

Figure 16. *Xerophyllum asphodeloides* (turkeybeard) habitat suitability map for the entire forest. (Basemap source: ESRI)

The final model shows the most likely suitable areas in Talladega National Forest as it represents areas where all models agree (Figure 16, Figure 17, Figure 18). Based on the final suitability prediction map, the most highly suitable areas are mostly located in the same geographical areas with the already known populations (Figure 17). Most of the probably suitable locations are also located on the north side of the national forest. Represented as a suitable area in Figure 14, Figure 15, Figure 16 include already known turkeybeard locations and new predicted populations.

Across all 4 methods, the most highly contributing (MaxEnt, RG) and significant (GLM, GAM) variables were ground slope and Bio4 (Temperature seasonality). Otherwise, the composition of each model's significant or highest contributing variables was different.

After generating all different prediction maps, the raster layers of the predictions for all four models' prediction maps were exported from R Studio and their accuracy checked in ArcGIS. Based on the final accuracy checking for the final ensemble map, 59% of the most probable suitable areas (most potential suitable areas from the ensemble model) overlapped with the already known occurrence locations (Figure 17). For the 4 different models, the highest accuracy was found 100% for the random forest model. The MaxEnt model prediction overlapped 77.3%, and the generalized linear model overlapped 68%. The lowest percentage of overlapping was found at 59.1% for the generalized linear model. Based on this final accuracy test, most areas predicted suitable for all models and ensemble model coincided with the already known occurrence polygons.

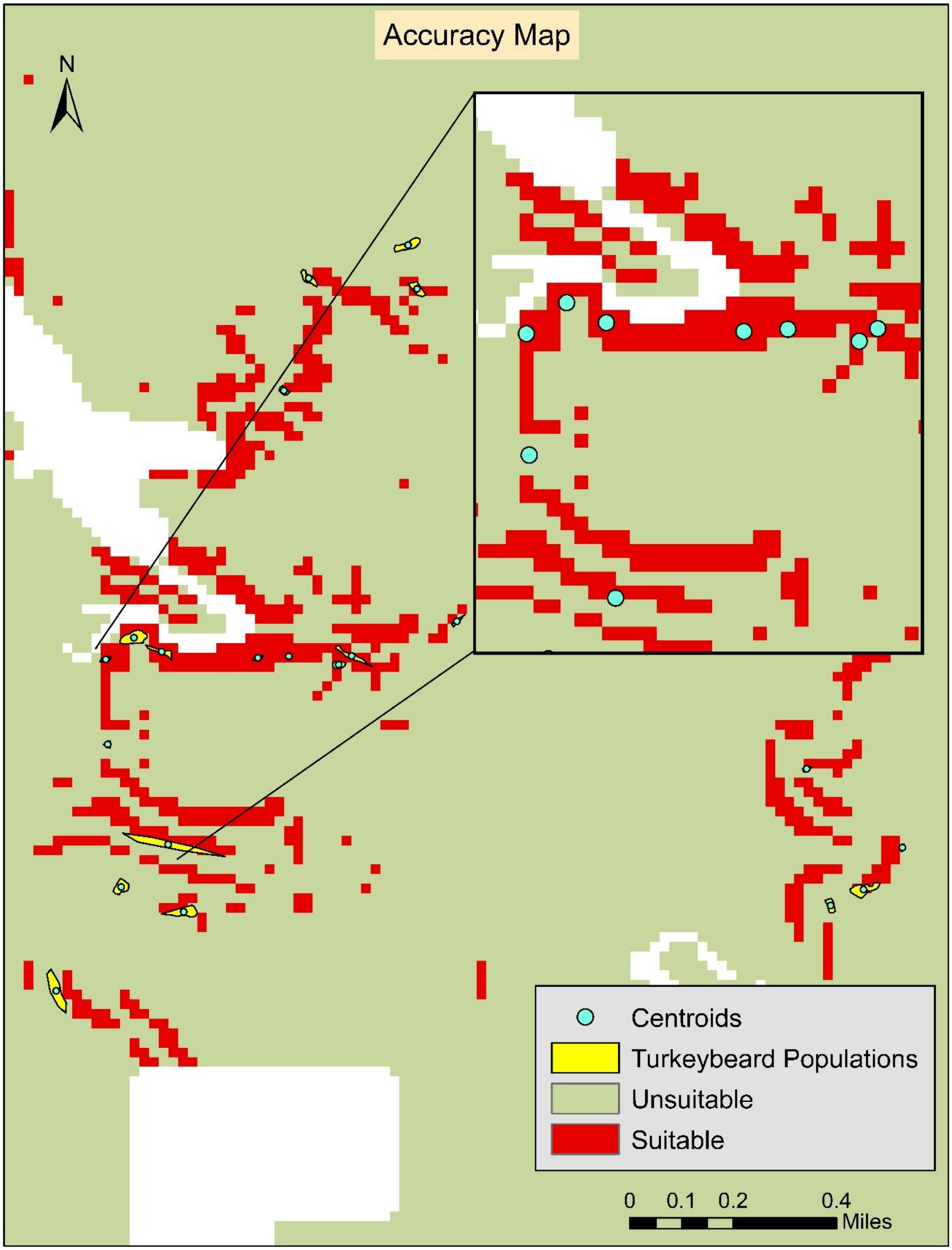

Figure 17. Accuracy map between already known turkeybeard locations and predictions.

CHAPTER 4

DISCUSSION

Habitat suitability modeling is important for understanding potential distribution areas for certain plant species to inform forest management and conservation practices. Habitat suitability modeling also helps determine relationships between environmental factors and species occurrences. In this study to understand the potential distribution of *X. asphodeloides* four common and successful modeling methods (MaxEnt, Generalized Linear Models, Generalized Additive Model, Random Forest) were used. This study is the first study of *X. asphodeloides* to identify the potential habitat suitability using multiple robust species distribution modeling methods. For the modeling effort, soil data, elevation, slope, aspect, and climate data (in total 28 variables) were used along with presence-only data of *X. asphodeloides*.

One of the most important limitations of this study was the limited presence data for the species (only 22 sites). Turkeybeard species are located in 10 states, but each state has very low occurrences because of the species rareness. Alabama has no information about the species occurrences. Only 22 occurrence locations were provided from the Talladega National Forest Management to generate the potentially suitable areas of the species. Based on the literature, sample size effects become less critical above 50 presences, even though some studies claim that some complex methods (MaxEnt) can cope with a small sample size (Stockwell et al., 2002; Pearson et al., 2007; Thibaud et al., 2014; Guisan et al., 2017). Guisan et al. (2017) also advise using 20 to 50 presence locations. Hernandez et al. (2006) also claimed that MaxEnt modeling can produce useful results with a sample size as small as 5 occurrences; however, the sample size has a profound impaction the predicted probability (Li and Ding, 2016).

Although there is this limitation of the present occurrences, this study results show that our AUC scores for each modeling approach were good. The results show using a small sample size for MaxEnt, GLM, GAM, and RF models can provide adequate prediction results. Based on the literature, the results of MaxEnt and RF were expected but GLM and GAM AUC scores were surprising. But as expected, the predictions of the suitable areas across the whole of the Talladega National Forest were very low; MaxEnt is 0.77%, GLM is 0.92%, GAM is 0.56%, and RF is 0.12% of the entire National Forest land base. The potentially suitable area that represents where all models agree is 0.09% of the entire national forest land base or 79 ha.

The other limitation of this study was the geographical distribution of the occurrences. When building a model, it is critical to incorporate as many geographically diverse samples as possible (Li and Ding, 2016). The occurrence sites are essentially located in the same geographical places. Because of that, the models predicted a small area of suitable habitat although the area under the curve (AUC) scores were high. These predictions show potentially suitable areas based on the species rareness and the effects of the lack of geographical diversity of the already known occurrence points.

In this study, the sensitivity-specificity approach was initially conducted to examine the appropriate thresholds per model, as recommended by Liu et al. (2005). However, these thresholds were replaced by a universal 0.5 threshold which is widely used in ecology (Manel et al., 1999; Bailey et al. 2002). This decision was made for several reasons. First, because some of the thresholds were low (such as 0.148 for the GLM), this would have created presence data for anything above this low threshold, which would have rendered low accuracy maps. Second, using the same threshold across all four models was necessary to avoid low probability and accuracy problems when creating a final combined habitat suitability map. Third, given the low prevalence of the species in TNF, a map showing the highest likelihood of suitability may provide the best information suggesting new areas to search for, or transplant the species.

The first study about the habitat prediction of the turkeybeard was conducted by Bourg et al. (2005). They conducted their research in Virginia using 24 occurrences population of the turkeybeard species. For their analyses, they used elevation, slope, aspect, forest type, six soil variables (composition, depth, drainage, pH, fertility, and water capacity), and fire frequency variables. Their models did not include the climate variables, but their models included the fire frequency variables, unlike this research. They used the classification and regression tree (CART) modeling using GIS, and generally, their modeling effort was successful at defining suitable habitats for turkeybeard species like this research. They found after their classification tree modeling that 4 variables (elevation, slope, forest type, fire frequency) were major determinants for estimating turkeybeard distribution. Unlike their study, my study concerning turkeybeard habitat predictions found the most important variables were slope and temperature seasonality; however, fire history was not used here. Based on the previous research and this research, the mutual environment variables are only ground slope for the distribution of the turkeybeard.

Several studies found climate change and soil important for the plant abundance and vegetation growth and distribution of the plant species (Walthert and Meier, 2017; Chakraborty et al. 2019; Buri et al. 2020). Based on this guidance, in this study, I used climate and soil data to generate potentially suitable areas for the turkeybeard. Chauvier et al. (2021) found in their study that the most important environmental factors for the distribution of the plant species involved the climate. In my work, these climate variables were also found effective, but soil variables were not found significant effects. This can be caused by the known locations not being geographically diverse. For further study, if geographically diverse occurrence populations were included in the model, it will be clearer to see the effects of the soil components for the species distribution.

Besides the importance of the environmental distribution of the turkeybeard species, future studies can also focus on the social influence of the species suitability. For instance, in the field observations, the turkeybeard species was located at the top of the hills and within some sheltering areas, so maybe its location can be related to ancient Native American peoples who lived in places that have a relation with these occurrences. Because I assumed that maybe they protected this species around their living areas to use turkeybeard for building materials and fire-starting material. There are no certain findings of this hypothesis, but these research suggestions can be conducted by anthropologists to see whether any social trigger for the distribution of the turkeybeard species might be plausible.

For further studies, new occurrence locations should be added to the modeling processes and the new occurrences should be geographically diverse. Based on this study experience, further studies about the turkeybeard species can be conducted using Maxent, RF, GAM, and GLM. If the models include more occurrences and more geographically diverse populations, it can generate a more accurate estimation of the potentially suitable areas. *Xerophyllum asphodeloides* is also one of the few known fire-adapted forest understory plant species and fire has been shown to have a significant positive effect on the reproductive performance of turkeybeard (Bourg et al., 2015). Based on this importance, further studies of turkeybeard habitat suitability modeling might also include some fire frequency or fire history variables to conduct suitability models.

This study also shows that it can be possible to use these models to predict the suitable habitat for rare and limited known species. This study can be conducted for the other rare, threatened, or endangered species distribution studies. Research findings also profit from the knowledge of this less-studied species to understand turkeybeard better for future management. The knowledge gained from this study can be used in future research about the turkeybeard species and the findings can be used the further field surveys to find new populations of the species. This study can inform the forest managers about future actions and future planning processes with respect to sustainable environmental practices and approaches. They can use this model to identify the distribution of the important species in their planning landscape. It will be also useful to make a decision under climate change effects and to identify the future distribution of the species under climate change. They can inform this modeling approach for their future planning and decision-making process. This model also helpful for the experts or managers to obtain more knowledge about the species habitat requirements. This study can be also important for conservation plans or further distribution (through human intervention) of the turkeybeard species. This preliminary modeling can be useful to find new populations and also can be useful to understand the relationship between turkeybeard and its habitats.

CHAPTER 5

CONCLUSION

This study represents the first effort to model potentially suitable habitat for the rare and threatened *X. asphodeloides* in Alabama and the Talladega National Forest. This study's findings show that the distribution of *X. asphodeloides* is highly related to ground slope and temperature seasonality. Excluding known locations, new areas that models indicated as potentially suitable for *X. asphodeloides* were very small in size and extent. However, based on the very low known presence of the species in the study area, this was expected. The given soil data was found not very effective based on the already known occurrence locations because the turkeybeard species exist in very similar soil characteristics; their geographical diverse is very low. To improve models, surveys for *X. asphodeloides* populations across a larger range of soil types throughout the Talladega National Forest would be useful. Prediction maps from this study can be used to guide future field surveys and improve the detection of unknown populations in the study area. For this rare plant species field survey, an ecologically informed meander search technique can be used, and this technique can be the more appropriate approach. This study also important to further the forest management planning processes to help schedule activities and maybe help develop conservation plans for this rare and threatened species. There are no apparent protection plans for turkeybeard species in the Talladega National Forest management plan, but they include protection plans in the future after getting more occurrence locations. In the management plan, they can reduce the harvesting activities on the already know locations until they get more knowledge about the turkeybeard and its habitat requirements. Knowledge gained from this research can be used to establish and implement habitat management strategies across the Talladega National Forest. This rare species can be used to the understand impact of climate change with modeling under the climate change scenarios. Incorporating habitat suitability modeling into the management plan will allow for policies and programs to be put into action to maintain and restore sustainable environmental practices. This research can therefore be used to inform managers about appropriate natural resource practices to maintain and restore sustainable forest management regimes with a conservation perspective. In sum, this research describes the development and application of a suitable ensemble habitat map model for the Talladega National Forest.

REFERENCES

Bahn, V., and B.J. McGill. 2013. Testing the predictive performance of distribution models. Oikos, 122(3): 321–331.

Bailey, S.A., R.H. Haines-Young, and C. Watkins. 2002. Species presence in fragmented landscapes: modelling of species requirements at the national level. Biological Conservation, 108(3): 307-316.

Bart, J., D.R. Petit, and G. Linscombe. 1984. Field evaluation of two models developed following habitat evaluation procedures. Transactions of the North American Wildlife and Natural Resource Conference, 49: 489–499.

Bender, L.C., G.J. Roloff, and J.B. Haufler. 1996. Evaluating confidence intervals for habitat suitability models. Wildlife Society Bulletin, 24(2): 347–352.

Benito, B.M., M.M. Martínez-Ortega, L.M. Muñoz, J. Lorite, and J. Peñas. 2009. Assessing extinction-risk of endangered plants using species distribution models: a case study of habitat depletion caused by the spread of greenhouses. Biodiversity and Conservation, 18(9): 2509-2520.

Booth, T.H. 2018. Species distribution modelling tools and databases to assist managing forests under climate change. Forest Ecology and Management, 430: 196–203.

Booth, T.H., H.A. Nix, M.F. Hutchinson, and T. Jovanic. 1988. Niche analysis and tree species introduction. Forest Ecology and Management, 23(1): 47–59.

Bourg, N.A., D.E. Gill, and W.J. McShea. 2015. Fire and canopy removal effects on demography and reproduction in turkeybeard (*Xerophyllum asphodeloides*), a fire-dependent temperate forest herb. Journal of Sustainable Forestry, 34(1-2): 71–104.

Bourg, N.A., W.J. McShea, and D.E. Gill. 2005. Putting a CART before the search: Successful habitat prediction for a rare forest herb. Ecology, 86(10): 2793–2804.

Brambilla, M., F. Casale, V. Bergero, G.M. Crovetto, R. Falco, I. Negri, P. Siccardi, and G. Bogliani. 2009. GIS-models work well, but are not enough: Habitat preferences of *Lanius collurio* at multiple levels and conservation implications. Biological Conservation, 142(10): 2033-2042.

Breiman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone. 1984. Classification and Regression Trees. Wadsworth International Group, Belmont, CA.

Breiman, L. 2001. Random forests. Machine Learning, 45(1): 5-32.

Buri, A., S. Grand, E. Yashiro, T. Adatte, J. E. Spangenberg, E. Pinto-Figueroa, E. Verrecchia, and A. Guisan. 2020. What are the most crucial soil variables for predicting the distribution of mountain plant species? A comprehensive study in the Swiss Alps. Journal of Biogeography, 47(5): 1143–1153.

Center for Plant Conservation. 2020. Plant Profile - *Xerophyllum asphodeloides*. Center for Plant Conservation, Escondido, CA. Retrieved from https://saveplants.org/national-collection/plant-search/plant-profile/?CPCNum=6634#DistOcc. Last assessed on July 16, 2020.

Chakraborty, D., R. Jandl, S. Kapeller, and S. Schueler. 2019. Disentangling the role of climate and soil on tree growth and its interaction with seed origin. Science of the Total Environment, 654: 393–401.

Chatterjee, S., and A.S. Hadi. 2006. Regression analysis by example. John Wiley & Sons, Hoboken, NJ.

Chauvier, Y., W. Thuiller, P. Brun, S. Lavergne, P. Descombes, D.N. Karger, J. Renaud, and N.E. Zimmermann. 2021. Influence of climate, soil, and land cover on plant species distribution in the European Alps. Ecological Monographs, 91(2): e01433.

Clark, J.D., and J.C. Lewis. 1983. A validity test of a habitat suitability index model for clapper rail. Proceedings of the Annual Conference of the Southeastern Association of Fish and Wildlife Agencies, 37: 95–102.

Cook, J.G., and L.L. Irwin. 1985. Validation and modification of a habitat suitability model for pronghorns. Wildlife Society Bulletin, 13(4): 440–448.

Di Biase, R.M., L. Fattorini, and M. Marchi. 2018. Statistical inferential techniques for approaching forest mapping. A review of methods. Annals of Silvicultural Research, 42(2): 46–58.

Edenius, L., and G. Mikusiński. 2006. Utility of habitat suitability models as biodiversity assessment tools in forest management. Scandinavian Journal of Forest Research, 21(S7): 62-72.

Fernald, M.L. 1950. Gray’s Manual of Botany (8th ed.). American Book Company, New York, NY.

Ferrarini, A., A. Selvaggi, T. Abeli, J.M. Alatalo, S. Orsenigo, R. Gentili, and G. Rossi. 2016. Planning for assisted colonization of plants in a warming world. Scientific Reports, 6(1): 1-6.

Fick, S.E. and R.J. Hijmans. 2017. WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. International Journal of Climatology, 37(12): 4302-4315.

Garzón, M.B., R. Blazek, M. Neteler, R.S. de Dios, H.S. Ollero, and C. Furlanello. 2006. Predicting habitat suitability with machine learning models: The potential area of *Pinus sylvestris* L. in the Iberian Peninsula. Ecological Modelling, 197(3-4): 383-393.

Giles, R.H., Jr. 1978. Wildlife Management. W.H. Freeman, San Francisco, CA.

Green, R.H. 1979. Sampling Design and Statistical Methods for Environmental Biologists. John Wiley & Sons, New York, NY.

Guisan, A., T.C. Edwards Jr., and T. Hastie. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling, 157(2-3): 89-100.

Guisan, A., A. Lehmann, S. Ferrier, M. Austin, J.M.C. Overton, R. Aspinall, and T. Hastie. 2006. Making better biogeographical predictions of species' distributions. Journal of Applied Ecology, 43(3): 386-392.

Guisan, A., W. Thuiller, and N.E. Zimmermann. 2017. Habitat Suitability and Distribution Models: with Applications in R. Cambridge University Press, Cambridge, UK.

Guisan, A., and N.E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecological Modelling, 135(2-3): 147–186.

Hair, J.F., R.E. Anderson, R.L. Tatham, and W.C. Black. 1998. Multivariate Data Analysis (5th ed.). Prentice Hall, Upper Saddle River, NJ.

Harrell, F.E. 2001. Resampling, validating, describing, and simplifying the model. In F.E. Harrell (Ed). Regression Modeling Strategies (pp. 87-103). Springer, New York, NY.

Hastie, T., and R. Tibshirani. 1990. Exploring the nature of covariate effects in the proportional hazards model. Biometrics, 46(4): 1005-1016.

Heikkinen, R.K., M. Marmion, and M. Luoto. 2012. Does the interpolation accuracy of species distribution models come at the expense of transferability? Ecography, 35(3): 276–288.

Hernandez, P.A., C.H. Graham, L.L. Master, and D.L. Albert. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography, 29(5): 773-785.

Hijmans, R.J., and C.H. Graham. 2006. The ability of climate envelope models to predict the effect of climate change on species distributions. Global Change Biology, 12(12): 2272–2281.

Hirzel, A.H., V. Helfer, and F. Metral. 2001. Assessing habitat-suitability models with a virtual species. Ecological Modelling, 145(2-3): 111-121.

Imam, E., S.P.S. Kushwaha, and A. Singh. 2009. Evaluation of suitable tiger habitat in Chandoli National Park, India, using spatial modelling of environmental variables. Ecological Modelling, 220(24): 3621-3629.

Keener, B.R., A.R. Diamond, L.J. Davenport, P.G. Davison, S.L. Ginzbarg, C.J. Hansen, C.S. Major, D.D. Spaulding, J.K. Triplett, and M. Woods. 2020. Alabama Plant Atlas. University of West Alabama, Livingston, AL. Retrieved from http://floraofalabama.org/Plant.aspx?id=4502. Last assessed on July 16, 2020.

Kliskey, A.D., E.C. Lofroth, W.A. Thompson, S. Brown, and H. Schreier. 1999. Simulating and evaluating alternative resource-use strategies using GIS-based habitat suitability indices. Landscape and Urban Planning, 45(4): 163-175.

Latif, Q.S., V.A. Saab, J.G. Dudley, and J.P. Hollenbeck. 2013. Ensemble modeling to predict habitat suitability for a large-scale disturbance specialist. Ecology and Evolution, 3(13): 4348–4364.

Latif, Q.S., V.A. Saab, K. Mellen-Mclean, and J.G. Dudley. 2015. Evaluating habitat suitability models for nesting white-headed woodpeckers in unburned forest. Journal of Wildlife Management, 79(2): 263-273.

Lee, T., P. Bettinger, C.J. Cieszewski, and A.R. Gutierrez Garzon. 2020. The applicability of recreation-grade GNSS receiver (GPS watch, Suunto Ambit Peak 3) in a forested and an open area compared to a mapping-grade receiver (Trimble Juno T41). PLoS ONE, 15(4): e0231532.

Lele, S.R., E.H. Merrill, J. Keim, and M.S. Boyce. 2013. Selection, use, choice and occupancy: clarifying concepts in resource selection studies. Journal of Animal Ecology, 82(6): 1183–1191.

Li, Y., and C. Ding. 2016. Effects of sample size, sample accuracy and environmental variables on predictive performance of MaxEnt model. Polish Journal of Ecology, 64(3): 303-312.

Liu, C., P.M. Berry, T.P. Dawson, and R.G. Pearson. 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography, 28(3): 385-393.

Liu, S., Y. Lin, Y. Zhang, Z. Guo, L. Zhang, C. Li, and J. Wang. 2011. Landscape ecology contributions to forestry and forest management in China: progresses and research needs. In C. Li, R. Lafortezza, and J. Chen (Eds). Landscape Ecology in Forest Management and Conservation (pp. 22-45). Springer, Berlin, Heidelberg.

Manel, S., J.M. Dias, and S.J. Ormerod. 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan River bird. Ecological Modelling, 120(2-3): 337-347.

McCullagh, P., and J.A. Nelder. 1989. Generalized Linear Models. Chapman and Hall, London.

McDonald, T.L. 2013. The point process use-availability or presence-only likelihood and comments on analysis. Journal of Animal Ecology, 82(6): 1174–1182.

Menard, S. 2002. Applied Logistic Regression Analysis (2th ed.). Sage Publishing, Thousand Oaks, CA.

Morrison, M.L. 2012. The habitat sampling and analysis paradigm has limited value in animal conservation: a prequel. Journal of Wildlife Management, 76(3): 438–450.

Nuttall, T. 1818. Genera of North American plants and a catalogue of the species, to the year 1817. Vol. 1, Printed for the author by D. Heartt, 1818. Sabin Americana: History of the Americas, 1500-1926.

Pearson, R.G., C.J. Raxworthy, M. Nakamura, and A. Townsend Peterson. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. Journal of Biogeography, 34(1): 102-117.

Pecchi, M., M. Marchi, V. Burton, F. Giannetti, M. Moriondo, I. Bernetti, M. Bindi, and G. Chirici. 2019a. Species distribution modelling to support forest management. A literature review. Ecological Modelling, 411: Article 108817.

Pecchi, M., M. Marchi, F. Giannetti, I. Bernetti, M. Bindi, M. Moriondo, F. Maselli, L. Fibbi, P. Corona, D. Travaglini, and G. Chirici. 2019b. Reviewing climatic traits for the main forest tree species in Italy. iForest, 12(2): 173–180.

Phillips, S.J., R.P. Anderson, and R.E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling, 190(3-4): 231-259.

Phillips, S.J., and M. Dudík. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography, 31(2): 161-175.

Rehfeldt, G.E., J.J. Worrall, S.B. Marchetti, and N.L. Crookston. 2014. Adapting forest management to climate change using bioclimate models with topographic drivers. Forestry, 88(5): 528–539.

Roloff, G.J., and B.J. Kernohan. 1999. Evaluating reliability of habitat suitability index models. Wildlife Society Bulletin, 27(4): 973-985.

Rosenblatt, F. 1958. The perceptron: a probabilistic model for information storage and organization in the brain. Psychological Review, 65(6): 386-408.

Rotenberry, J.T., K.L. Preston, and S.T. Knick. 2006. GIS-based niche modeling for mapping species' habitat. Ecology, 87(6): 1458–1464.

Shataee, S., S. Kalbi, A. Fallah, and D. Pelz. 2012. Forest attribute imputation using machine-learning methods and ASTER data: Comparison of k-NN, SVR and random forest regression algorithms. International Journal of Remote Sensing, 33(19): 6254-6280.

Stober, J., K. Merry, and P. Bettinger. 2020. Analysis of fire frequency on the Talladega National Forest, USA, 1998-2018. International Journal of Wildland Fire, 29(10): 919-925.

Stockwell, D.R., and A.T. Peterson. 2002. Effects of sample size on accuracy of species distribution models. Ecological Modelling, 148(1): 1-13.

Swets, J.A. 1988. Measuring the accuracy of diagnostic systems. Science, 240(4857): 1285-1293.

Thibaud, E., B. Petitpierre, O. Broennimann, A.C. Davison, and A. Guisan. 2014. Measuring the relative effect of factors affecting species distribution model predictions. Methods in Ecology and Evolution, 5(9): 947-955.

Thuiller, W., L. Brotons, M.B. Araujo, and S. Lavorel. 2004. Effects of restricting environmental range of data to project current and future species distributions. Ecography, 27(2): 165–172.

Thuiller, W., and T. Münkemüller. 2010. Habitat suitability modeling. In A.P. Moeller, W. Fielder, and P. Berthold (Eds.). Effects of Climate Change on Birds (pp. 77-85). Oxford University Press, Oxford, UK.

University of Alabama Department of Geography. 2020a. Alabama Ecoregions. Retrieved from http://alabamamaps.ua.edu/Interactive%20Maps/Physical/Ecoregions.html. Last assessed on December 6, 2020.

University of Alabama Department of Geography. 2020b. Alabama Historical Climate Data. Retrieved from http://alabamamaps.ua.edu/Interactive%20Maps/Climate/WeatherStations.html. Last assessed on January 28, 2021.

United States Department of Agriculture. 2012. Establishment and modification of national forest boundaries and national grasslands. A chronological record 1891-2012. U.S. Department of Agriculture, Forest Service, Lands and Realty Management Staff, Washington, D.C. FS-612.

United States Department of Agriculture. 2014. National Elevation Dataset 30 m. United States Department of Agriculture, Natural Resources Conservation Service. Washington, D.C. Available at http://datagateway.nrcs.usda.gov/.

United States Department of Agriculture. 2019. SSURGO-Web Soil Survey 3.3. Washington, D.C. Available at https://websoilsurvey.sc.egov.usda.gov/app/homepage.htm.

United States Department of Agriculture. 2020a. Plant Profile. United States Department of Agriculture, Forest Service, Washington, D.C. Retrieved from https://plants.usda.gov/core/profile?symbol=XEAS. Last accessed on July 16, 2020.

United States Department of Agriculture. 2020b. Forest Service Home. Forest Management. About Us. United States Department of Agriculture, Forest Service, Washington, D.C. Retrieved from https://www.fs.fed.us/forestmanagement/aboutus/ecoperspective.shtml#:~:text=The%20overriding%20objective%20of%20the,of%20forests%20through%20their%20actions. Last accessed on July 17, 2020.

United States Fish and Wildlife Service. 1981. Standards for the development of habitat suitability index models. United States Fish and Wildlife Service, Washington, D.C. Release No. 1-81,103 ESM.

United States Forest Service. 2004. Revised land and resource management plan, national forests in Alabama. USDA Forest Service, Southern Region, Management Bulletin R8-MB 112A, Atlanta, GA.

Verner, J., M.L. Morrison, and C.J. Ralph. 1986. Wildlife 2000, Modeling Habitat Relationships of Terrestrial Vertebrates. University of Wisconsin Press, Madison, WI.

Walthert, L., and E. S. Meier. 2017. Tree species distribution in temperate forests is more influenced by soil than by climate. Ecology and Evolution, 7(22): 9473–9484.

Wei, B., R. Wang, K. Hou, X. Wang, and W. Wu. 2018. Predicting the current and future cultivation regions of *Carthamus tinctorius* L. using MaxEnt model under climate change in China. Global Ecology and Conservation, 16: Article e00477.

Wenger, S.J., and J.D. Olden. 2012. Assessing transferability of ecological models: an underappreciated aspect of statistical validation. Methods in Ecology and Evolution, 3(2): 260–267.

Wiens, J.A. 1986. Spatial and temporal variation in studies of shrub steppe birds. In J. Diamond and T.J. Case, (Eds). Community Ecology (pp. 154-172). Harper and Row, New York, NY.

Wiens, J.A. 2002. Predicting species occurrences: Progress, problems, and prospects. In J.M. Scott, P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall, and F.B. Samson (Eds.). Predicting Species Occurrence: Issues of Accuracy and Scale (pp. 739-749). Island Press, Washington, D.C.

Yee, T.W., and N.D. Mitchell. 1991. Generalized additive models in plant ecology. Journal of Vegetation Science, 2(5): 587-602.

APPENDIX A

Detailed Explanation of Bioclimatic Variables

Code	Explanation
BIO1	Annual Mean Temperature
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3	Isothermality (BIO2/BIO7) (×100)
BIO4	Temperature Seasonality (standard deviation ×100)
BIO5	Max Temperature of Warmest Month
BIO6	Min Temperature of Coldest Month
BIO7	Temperature Annual Range (BIO5-BIO6)
BIO8	Mean Temperature of Wettest Quarter
BIO9	Mean Temperature of Driest Quarter
BIO10	Mean Temperature of Warmest Quarter
BIO11	Mean Temperature of Coldest Quarter
BIO12	Annual Precipitation
BIO13	Precipitation of Wettest Month
BIO14	Precipitation of Driest Month
BIO15	Precipitation Seasonality (Coefficient of Variation)
BIO16	Precipitation of Wettest Quarter
BIO17	Precipitation of Driest Quarter
BIO18	Precipitation of Warmest Quarter
BIO19	Precipitation of Coldest Quarter

APPENDIX B

The Correlation Matrix for All Environmental Variables