Assessing the Transferability of a Species Distribution Model for Predicting the Distribution of Invasive Cogongrass in Alabama

by

Rachel Eagle Shanks

A Thesis Presented to the
Faculty of the USC Graduate School
University of Southern California
In Partial Fulfillment of the
Requirements for the Degree
Master of Science
(Geographic Information Science and Technology)

December 2019

To my mother, Angela Faye Eagle.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I am grateful to the faculty and staff of the Spatial Sciences Institute at the University of Southern California for the excellent program and exceptional professors and cohorts. I am especially grateful to my advisor, Dr. Karen Kemp for her guidance and encouragement through this process and ever-present patience as I worked through the challenges of life + graduate school. I am grateful for the time and experience offered to me by my committee members, Dr. Su Jin Lee and Dr. Katsuhiko Oda. This journey has been a long one, but one that has shaped my life and career for the better.

I would also like to thank the Alabama Forestry Commission for providing the full cogongrass presence point location dataset for use in this analysis as well as for guiding me toward valuable contacts, resources, and literature used in this study. I would like to thank my employer, Silvics Solutions, for providing the roads datasets used in this analysis and for being my sounding board when ArcGIS baffled my sometimes less than analytical brain. I would also like to send out special thanks to my manager and mentor, Greg Triplett, for allowing me the time to focus effort on this project when timelines required, for his patience and understanding when my priorities shifted, and for picking up my slack when I had to prioritize school over work.

Finally, I would like to thank my family without whom I would not have been able to juggle work, parenting, school, and life in general. Through all of the missed visits, distracted phone calls, and general negligence of my responsibilities as a mother, daughter, sister, and friend, we somehow came out of this still speaking to one another! I love you all and could not have completed this journey without your understanding and support.

List of Abbreviations

ASCII	American Standard Code for Information Interchange
AUC	Area Under the Curve
CSV	Comma-Separated Values
DEM	Digital Elevation Model
GAP	Gap Analysis Project
GIS	Geographic information system
GISci	Geographic information science
GIST	Geographical Information Science and Technology
MSA	Modeled Study Area
PRISM	Parameter-elevation Regressions on Independent Slopes Model
ROC	Receiver Operating Characteristic
SDM	Species Distribution Model
SSI	Spatial Sciences Institute
TSS	True Skill Statistic
USC	University of Southern California
USFS	United States Forest Service
USGS	United States Geological Survey

Abstract

As of April 19th, 2018, there were 34,771 verified locations of cogongrass (*Imperata cylindrica* (L.) Beauv.) infestations within the state of Alabama. Cogongrass is a highly invasive non-native species of rhizomatous grass that is considered one of the ten worst weeds worldwide. This highly invasive and environmentally destructive species has caused significant damage throughout its current distribution and efforts to control and eradicate the threat have been underway for almost a decade. This study utilized the Maximum Entropy (Maxent) model to predict the location of invasive cogongrass within the state of Alabama. The model developed using the presence locations and environmental data for the Model Study Area, one Alabama Forest Commission (AFC) Work Unit, was applied to two additional AFC Work Units to test transferability of the model to areas of similar and dissimilar ecological and geographic makeup. The Model Study Area's Maxent model resulted in an acceptable AUC (0.725 with sd = 0.0010) and fair TSS score (0.4087) with a test omission rate of 0.0832. Transferability test results differed between the two test areas. Using the Model Study Area's model on Test Area 1, an area similar in most aspects to the Model Study Area, resulted in an AUC of 0.746 with a standard deviation of 0.002, a TSS score of 0.3944 and a test omission rate of 0.0807. These results indicated that the original model was sufficiently transferable to the similar Test Area 1. Test Area 2 was dissimilar from the Model Study Area in most environmental covariates as well as number of verified presence point locations. Applying the model to Test Area 2 resulted in an AUC of 0.846 with a standard deviation of 0.017, a TSS score of 0.2377 and a test omission rate of 0.2941. These results suggest the need for some concern about the suitability of the transferred model to Test Area 2.

Chapter 1 Introduction

Imperata cylindrica (L.) Beauv., commonly known as cogongrass (Figure 1) is a highly invasive and environmentally destructive non-native species with serious biological, environmental, and economic impacts to the Southeastern United States. In fact, as of October 22nd, 2018, the U.S. Forest Service website lists cogongrass as "one of the 10 worst weeds worldwide and a pest in 73 countries." Cogongrass, like most non-native invasive species, can become an agent of change in the ecosystem within which it becomes established. As an agent of change, the species can have a deleterious effect on native biodiversity (McNeely 2001).

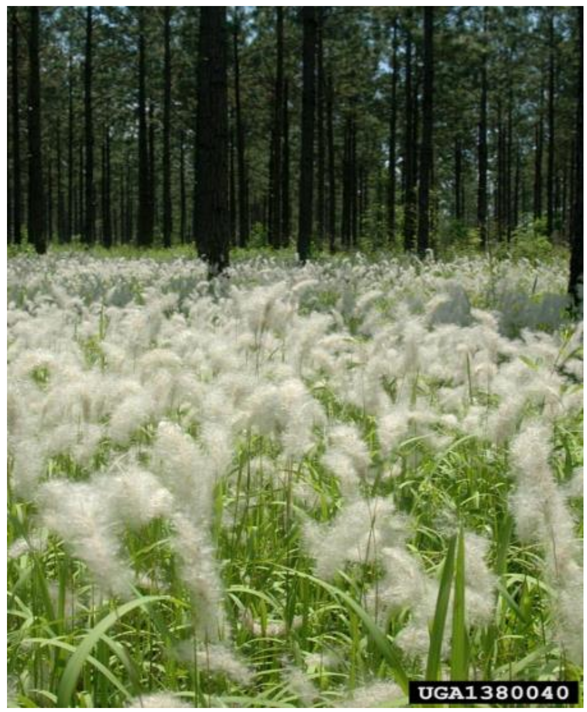


Figure 1: Image of Imperata cylindrica (L.) Beauv. in bloom in a pine plantation. Image curtesy of Chris Evans, University of Illinois with permission via bugwood.org.

In an effort to better understand the distribution and potential infestation threat of invasive species, ecologist use tools such as species distribution models (SDM) to assist in their understanding of the potential species spread and to plan for appropriate management actions related to the species being studied. The Maximum Entropy Model (Maxent) is a SDM which is commonly used by ecologists to study the current and predicted future distribution of a species given presence-only datasets. The use of this model helps researchers better predict infestation points based on environmental factors and assist in their efforts to eliminate this blight by guiding eradication funds to appropriate areas of high risks for infestation. Limited funding necessitates that eradication efforts must be focused on areas that respond best to treatment to ensure maximum benefit to the environment, community, and rural economy. In this analysis, Maxent was used to model the predicted potential distribution of cogongrass infestation given suitable conditions within Alabama Forestry Commission's (AFC) Work Unit 11, which in this document is referred to as the Model Study Area. The resultant model was then transferred to two other study areas to test model transferability for the species and to theorize the potential for model transferability across the state. The two transferability test areas were selected so that Test Area 1 was highly similar in ecological niche and number of verified infestation point locations to the Model Study Area and Test Area 2 is dissimilar. These study locations are shown in Figure 2.

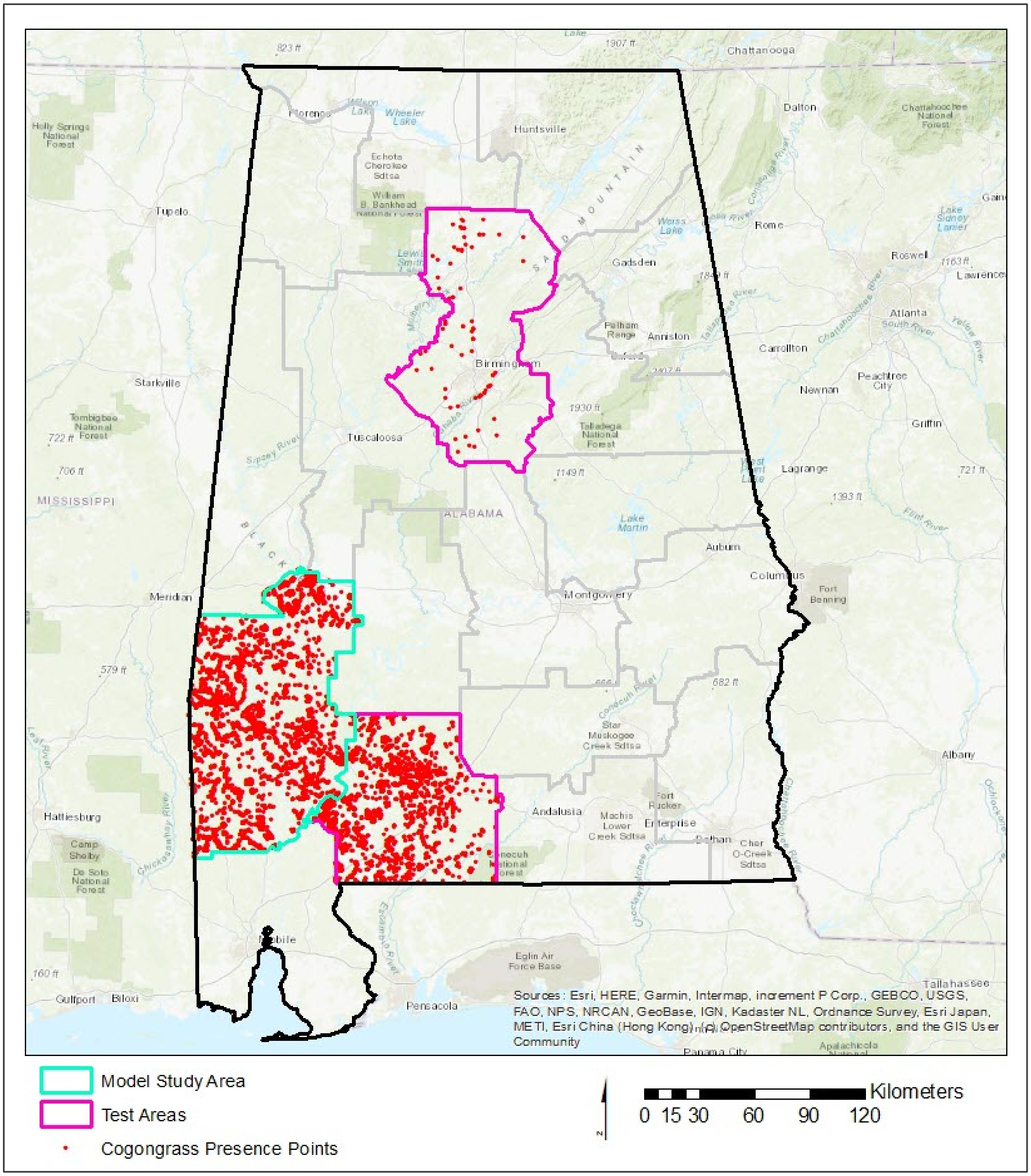


Figure 2: Map of cogongrass infestation presence point locations with the Model Study Area and two transferability test areas defined. The Model Study Area is outlined in cyan and the two Test Areas are outlined in Magenta.

1.1. Cogongrass

Imperata cylindrica (L.) Beauv. (cogongrass), is a highly invasive non-native species of rhizomatous grass that was originally introduced in the southeastern United States accidentally in 1912 as packing material in shipping crates from Japan for imported goods at the Port of Mobile in Grand Bay Alabama (Tabor 1949; Tabor 1952; Dickens 1974; Dozier 1998; MacDonald 2004; Damghani 2013). The species was later intentionally introduced from the Philippines in Mississippi (Tabor 1949; Tabor 1952; Dickens 1974; Dozier 1998; Ervin and Holly 2011) and Florida in the 1920s and 1930s by the USDA as forage and for erosion control (USDA NRCS Plants Database). The var. rubra variety (a non-invasive ornamental cultivar) of *Imperata cylindrica* is still sold by the nursery industry in some states as an ornamental grass under the name Japanese Blood Grass, or Red Baron, (Dozier 1998; Missouribotanicalgarden.org, last accessed 11/4/2018), however all other varietals are listed as a Federal Noxious Weed under the Plant Protection Act, which limits its transport between states without an appropriate permit.

Currently the range of verified infestations of cogongrass within the continental United States spans from East Texas, Southeast to South Florida, and as far north as North Carolina, according to the Early Detection and Distribution Mapping System website developed by The University of Georgia – Center for Invasive Species and Ecosystem Health (EDDMapS 2019). Figure 3 shows a map of this distribution.

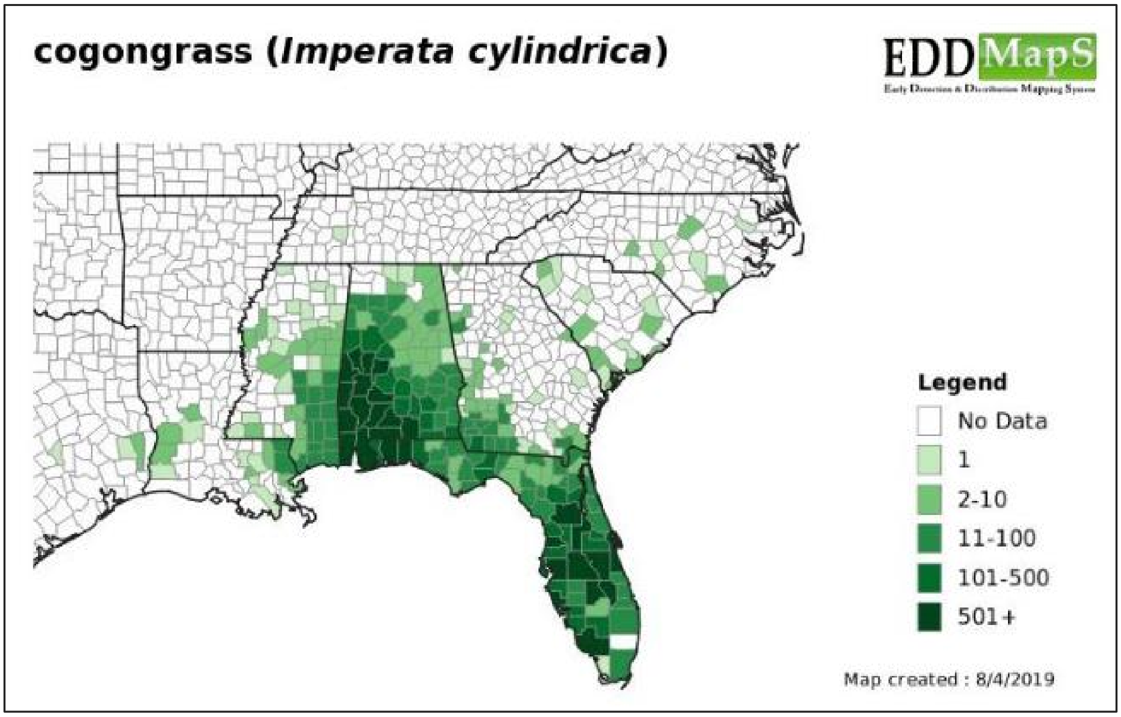


Figure 3: County level distribution with density of infestation points of cogongrass verified sites. Source: EDDMapS 2019

1.2. Research Goals

The research objectives of this study were two-fold. The first objective was to evaluate the fitness for use of Maxent in modeling the predicted potential distribution of cogongrass infestation given suitable conditions using the selected environmental covariates within the Model Study Area. The second objective was to test the transferability of that model to other study areas within the state of Alabama.

Alabama Forestry Commission Work Units were used to delineate the boundaries of study areas within this project. AFC Work Unit 11 was selected as the Model Study Area because the area contains a large verified point location dataset to use in the model (9242 points) and this AFC Work Unit contains the transferability study area from the Ervin and Holly (2011) study (Clarke County, Alabama) that initially sparked my interest in model transferability. Percent canopy cover was shown to be the most influential variable on the Ervin and Holly

Mississippi model, and therefore we hypothesize that percent canopy will have significant influence on the models produced in this study as well.

The two test study areas were selected based on their similarity and dissimilarity to the Model Study Area. It was hypothesized that Test Area 1, which is relatively similar in environmental covariate values to the Model Study Area, will have a similar model result to the Model Study Area. Further, it is hypothesized that Test Area 2, which is relatively dissimilar in environmental covariate values to the Model Study Area, will have dissimilar model results to the Model Study Area but will still produce an acceptable model.

The guiding motivation, beyond the desire to generate an appropriate model that is transferable across various areas of the state, is the hope that the resulting model and transferability tests will be useful in directing future survey efforts and funding decisions for implementing control and eradication measures against invasive cogongrass in the state of Alabama. Evaluating the model will help researchers better predict infestation points based on environmental factors used and assist in their efforts to eliminate this threat by guiding survey and eradication funds to appropriate areas of high risk for infestation. Invasive species management has been shown to be more effective when management activities occur in the early stages of infestation as attempted management of large, well-established colonies of invasives is difficult and cost prohibitive (Ervin and Holly 2011). Limited funding necessitates that eradication efforts must be focused on areas that respond best to treatment to ensure maximum benefit to the environment, community, and rural economy. The use of Maxent to facilitate targeted survey and eradication efforts is possible only if this type of SDM can be shown to be a useful tool in predicting the distribution and spread of this species and is transferable across the affected area. In addition, the results of this study can be used to further prompt research into this species as well as the use of Maxent in predicting species distribution.

1.3. Study Organization and Structure

This study was structured to first define an appropriate Maxent model for cogongrass in a specific area in the state of Alabama and then test the transferability of that model to other areas within the state. Figure 4 depicts the overall study workflow. First the project goals and species were defined. Then the model study area and study related questions were reviewed. These questions, and the answers to them, as gleaned from research, guided the definition of the species and environmentally appropriate datasets needed to complete the study. Once the required datasets were identified, the data was prepared for use by Maxent using Esri's ArcGIS 10.6 Desktop. The prepared data was then used within Maxent 3.4.1. A baseline model was trained utilizing all gathered datasets and all Maxent default values and then the model for the Model Study Area was tuned through iterative runs where maxent settings were modified and environmental covariates that were deemed to add little to no added value to the model were removed. A final model was created for the model study area and results were analyzed to verify fitness for use given the species and environmental extent of the study. The model produced for the Model Study Area was then used to test transferability to the two test areas.

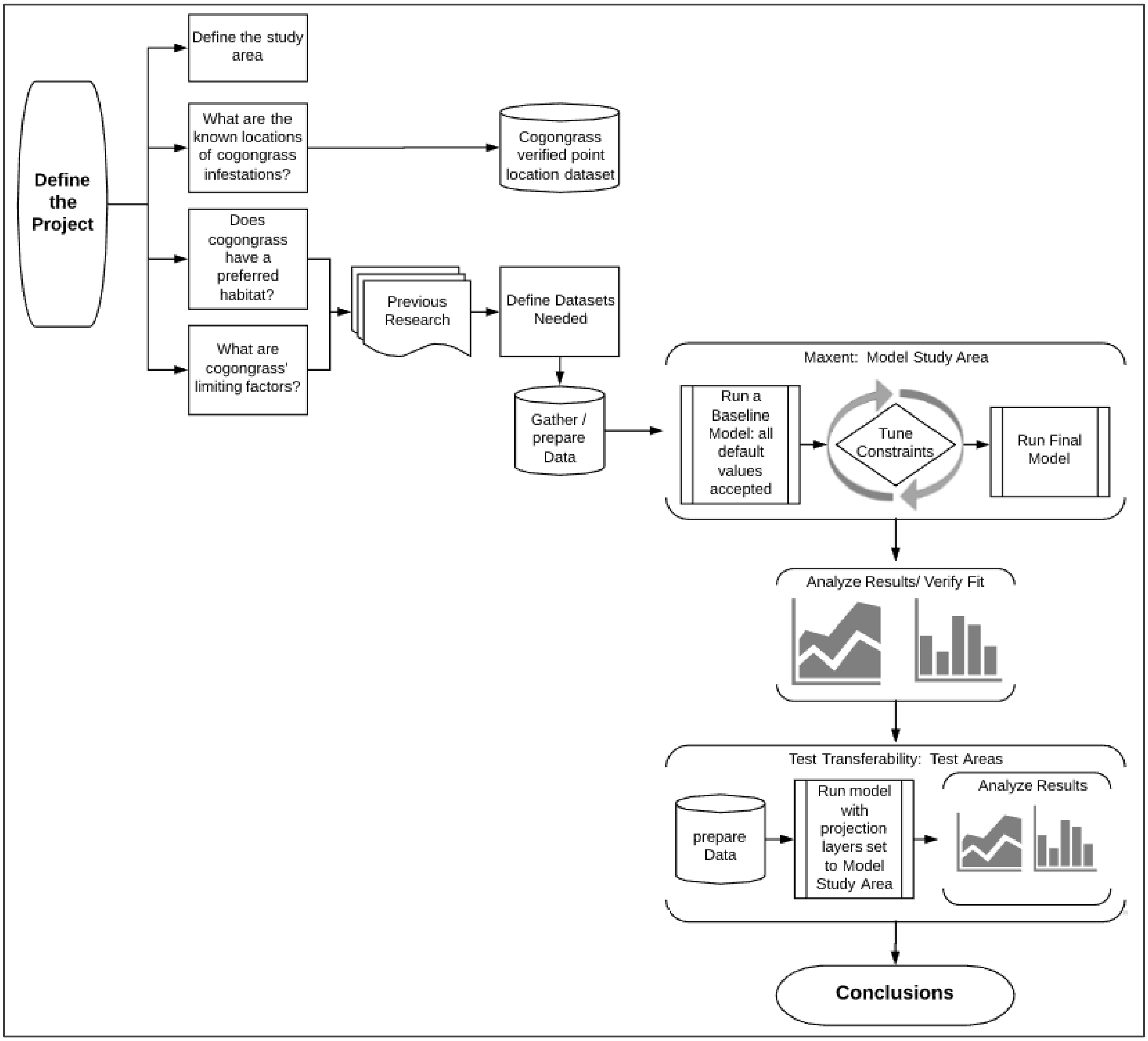


Figure 4: General structure and organization of the project.

The remainder of this document is broken into four additional chapters. The next chapter provides background context and additional information pertinent to the species being studied, the modeling method used, and research that guided the decisions made as this project progressed. Chapter 3 describes the data included in this study in detail, as well as the methods used to build the models generated by this project. Chapter 4 discusses the results of the models produced and specifically focuses on the key statistics for judging model fitness. Finally, Chapter

5 includes a discussion of the conclusions gleaned from this project and the models generated in the process of this study.

Chapter 2 Background

This chapter provides background context and additional information pertinent to the species being studied, the modeling method used, and research that guided the decisions made as this project progressed. This chapter begins by describing, in detail, the morphology and biological characteristics of the species and the habitat range in which it can grow. The chapter then continues by discussing Maxent as a tool for modeling and the tuning and testing of the output model. Finally studies pertinent to the decisions made in this study are reviewed.

2.1. Description of the Species

Cogongrass has many alternate common names throughout the world. It is also known as kunai grass, blady grass, japgrass, alang-alang, lalang grass, as well as many others and is often confused with Brazilian satintail (*Imperata brasiliensis*) which is a closely related species in the genus. Cogongrass is fast-growing and can spread by rhizomatous shoot up to $4m^2$ in as little as 11 weeks on productive sites (Dozier 1998; Wilcut et al. 1988a). The general biological characteristics of the species are defined in Table 1 below. The species is stemless, forming rigid leaf tufts developing directly from the rhizomes. Leaves can grow to 150cm in height and 4 to 10mm in width and have very sharp pointed apex. They have an off-center midrib that is white in appearance and has finely serrated margins (Estrada and Flory 2014; Dozier 1998). Cogongrass has a high rhizome to shoot ratio which increases its regenerative ability and exhibits allelopathic tendencies which inhibit growth of competing native species (MacDonald 2004).

Table 1: General biological characteristics of cogongrass

Biological Characteristics	Description
Reproduction	Vegetative and seed
Flower	Branched panicle with dense white fluffy spikelets growing 10-20cm long
Growth Structure	Stemless, grows in loose tufts with leaves emanating from rhizomes
Leaf Blade	Long and slender; 15-150cm tall; 4-10mm wide with off center white midrib
Root Structure	Rhizomes with attached dense fibrous root system

Rhizomes are the primary mechanism for local spread of the species once invasion has occurred. Rhizomes are aggressive, hardy, branched, and grow in dense clumps. These clumps form dense mono-species mats that impede the growth of other species that would otherwise utilize that environment. In a 1977 study by Lee et al., rhizome density was measured to be 89m (linear) per square meter of soil. Rhizome clumps restrict access to nutrients needed for native or commercial species to thrive, further harming both the biological and economic environments in which it is found. Rhizomes are whiteish in color with short, scaly nodes and sharp barbed tips that can penetrate the roots of other species (Dozier 1998). The morphology of the species is represented in Figure 5. It is important to note that buds do not form until the third or fourth leaf stage of the plant's life cycle. This is also when dense root development begins (Dozier 1998). The importance of this is due to planned timing of infestation eradication. For cogongrass, invasive species management has been shown to be more effective when management activities occur in the early stages of infestation as attempted management of large, well-established colonies of invasives is difficult and cost prohibitive (Ervin and Holly 2011). Therefore, eradication efforts will be less costly if management activities can be conducted in young plants.

Figure 5: Example morphology of cogongrass. Images reproduced by permission from bugwood.org.

Panicle flower heads are 5-20cm long and silvery-white. The panicle is fuzzy giving the flower a soft cottony look (EDDMapS 2019). Some studies suggest that flowering occurs generally after disturbance or stress but recent studies counter that thought and show that cogongrass produces an abundance of seed even without disturbance or stress and the seeds are easily distributed by wind. Each plant can produce up to 3000 seeds annually (MacDonald 2004; Dozier 1998; Wilcut et al. 1988a; Holm 1977). Cogongrass spreads locally via rhizome growth and long-distance via seed dispersal. Wilcut et al. (1988a) state the average flight of a one-seeded spikelet was 15m, and a 2011 study by Yager, Miller, and Jones measured the maximum flight distance for a spikelet, with seed removed, to be 37m in a pine-tallgrass environment. Studies have suggested that the West to East wind patterns along major roads and Interstate highways has created a dispersal route for cogongrass infestation by seed (Yager, Miller, and Jones 2011; MacDonald 2004; Wilcut et al. 1988a; Hubbard et al. 1944).

Cogongrass is a very hardy species and is tolerant of shade, high salinity, moisture and drought. The general habitat description of the species is defined in Table 2. Cogongrass grows in tropical and subtropical climates ranging in latitude from 45°N to 45°S. occurring in a wide range of ecological conditions (MacDonald 2004; Holm et al. 1977). Cogongrass thrives in minorly disturbed sites and non-disturbed rural sites but not heavily disturbed sites. The species has been shown to thrive along roadways, in pastures and mining sites, pine forests and other open areas, but does not thrive in areas of heavy cultivation and repeated tillage (Dozier 1998; Willard et al. 1990) Therefore one mechanism for control is repeat tillage treatments when infestation occurs on sites where tillage is possible, and it would be expected that cogongrass would not infest agricultural row crop sites where repeat heavy tillage occurs.

Table 2: General habitat description of cogongrass

Habitat	Description
Range	Tropical and subtropical climates (Latitudes 45ºN to 45ºS)
Site	Highly adaptable (occurs in a wide range of ecological conditions from xeric uplands to shaded mesic sites) • Degraded forests, roadsides, arable land, young plantations, sandhills, flatwoods, hardwood hammocks, grasslands, river margins, swamps, scrub, and wet pine savanna communities • Thrives in areas of minimal tillage and frequent burning • Tolerant of varied soil conditions including variations in fertility, organic matter and moisture • Grows best in relatively acidic soils (pH 4.7) • Relatively intolerant of shade
Rainfall	75 to 500cm
Elevation	Sea level to 2000m
Temperature	-4.5ºC or lower for more than 24 hours is lethal to rhizomes (however dense thickets can insulate themselves and may survive temperatures as low as -14ºC.

In 2009, the Alabama Forestry Commission received a three-year, 6.3-million-dollar grant from the American Recovery and Reinvestment Act to initiate a proactive, coordinated campaign to eradicate cogongrass in the state of Alabama. The Commission for the Campaign Against Cogongrass was formed to detect, map, and plan an effective program for the eventual eradication of cogongrass from the state. This grant was sufficient to get the initiative started; however more funding is necessary to win the war on cogongrass (Bargeron 2009).

According to the U.S. Fish and Wildlife Service Invasive Species website, invasive non-native species do not have the natural checks and balances that native species would have in an ecosystem. Therefore, when a non-native species is introduced to a new environment, it can become invasive if there are no natural elements to restrict its propagation. This invasion of the ecosystem by a non-native species can have deleterious effects on the ecosystem, the economy, and human health (US Fish and Wildlife Service last accessed 11/4/2018). Cogongrass has been an invasive non-native species in the southeastern United States since its introduction in 1912 and has been shown to have significant impacts both economically and environmentally in heavily infested areas. Figure 6 shows an aerial view of the impact of cogongrass infestation in a young pine plantation. Cogongrass is a highly adaptive invasive species with a broad tolerance to environmental and ecological conditions. Therefore, this species has the potential to adversely change the structure and diversity of environments in which it invades. Cogongrass has been linked to the reduction of native diversity and alteration of ecological processes within infested ecosystems, especially in fire-dependent communities (Lippincott 2000). This highly invasive and environmentally destructive species has caused significant damage throughout its current distribution and efforts to control and eradicate the threat have been underway for almost a decade.

Figure 6: Cogongrass infestation in a young pine plantation exhibiting its distinctive circular infestation pattern and severity of infestation at occurrence locations. Image courtesy of Greg Leach, International Paper via Bugwood.org.

Economic stressors resulting from the establishment of cogongrass include cost of eradication, impact of eradication efforts on native species and agricultural crops, financial loss due to disturbance, etc. (Hubbard et al. 1944; Soerjani 1970; Eussen et al. 1976; Daneshgar et al. 2008). Studies have also shown that cogongrass is particularly problematic in agricultural systems where the species directly competes with agricultural crops for both space and nutrients. This competition reduces crop yields and increases weed control costs (Ervin and Holly 2011; Akobundu and Ekeleme 2000; Terry et al. 1997). Cogongrass has been the subject of numerous and diverse studies throughout the Southeastern United States and therefore is a good candidate for studying the effectiveness of a Maximum Entropy model (Phillips et al. 2017) in predicting its distribution and testing transferability of the same.

2.2. Modeling with Maxent

SDMs are routinely used to predict the potential distribution of a species based on known point locations. The species distribution modeling used in this study is the maximum entropy method using Maximum Entropy Species Distribution Modeling (Maxent) Version 3.4.1 (Phillips et al. 2017). Maxent uses presence only species location points and environmental variables to develop probability models for the distribution of the species being modeled (Phillips et al. 2006; Ervin and Holly 2011; Elith et al. 2011). Presence locations are compared to the environment through the use of background points (Crall et al. 2013).

Maxent has been gaining in popularity and use in the fields of ecology and environmental sciences and has been shown to outperform other species distribution modeling methods in predictive accuracy (Merrow, Smith, and Silander 2013). Machine Learning models such as the Maxent model allow the model to "learn" from iterative model runs given a sample known dataset to train the model and a sample known dataset to test the model's understanding of the data and associated environmental layers.

Maxent is a machine learning model that is well suited for species distribution modeling (Phillips et al. 2017). Maxent models are good predictors of species distribution with limited datasets and work especially well with presence only data (Phillips et al. 2017; Ervin and Holly 2011; Elith et al. 2011). Before running the model, the biology and ecological niche of the species being studied must be closely examined to ensure that selected environmental variables align with the biological and environmental factors that influence the species distribution within the intended study area (Manzoor, Griffiths, and Lukac 2018). This species review and conscientious environmental data selection can be time consuming but is vital to the production of an informed SDM. The Maxent model allows for an understanding of which environmental variables are most important to the distribution of the species being modeled and gives a relatively unbiased prediction based on the constraints provided. The model requires a training dataset, a testing dataset, and environmental layers to act against the model as predictors (Merrow, Smith, and Silander 2013). The training and testing datasets are presence only data and can be subsets of the same original larger dataset.

As Maxent is a presence only modeling application, the pseudo-absence (background) points are generated by the model. It is important to note that some sampling bias may be introduced into the model if your species presence point distribution does not cover the entire range of your study's geographic extent. If this is the case, you can create a background file in ArcGIS to use within Maxent to limit where the model predicts background points so that it does not create background points outside of the extent of the presence data points. This can be done using the Create Minimum Bounding Geometry tool in ArcGIS then converting the output polygon to raster (.asc) format for use in Maxent. In this study, no minimum bounding geometry was needed as the presence points spanned the entire extent of the Model Study Area.

2.2.1. Model Tuning

In Maxent, model tuning is performed to optimize model complexity and fit. Tuning the model smooths the response curves to the specific environmental variables included in the model to reduce overfitting (Elith et al. 2011). Maxent provides default settings for parameters that were determined to be the average optimal values (Phillips and Dudık 2008), however, it is recommended that these settings be tuned for the specific species and region of study (Radosavljevic and Anderson 2014).

To assist with potential issues related to spatial autocorrelation, a baseline (neutral) model run can be performed with all parameters set to default. Based on the results of the baseline model run, the model settings can be tuned until an appropriate model output from the sample data is derived. Modifications to parameters, constraints, and environmental layers included should be based on research of similar studies (Merrow, Smith, and Silander 2013).

Regularization is an available parameter in Maxent that relaxes the environmental constraints so that the predictions do not have to fit the constraints exactly. This allows the model to ignore variables that don't impact the model and to determine the most impactful variables on the model output. Regularization protects against overfitting by affecting how closely the output distribution is fit to the provided presence data. To get a closer fit (more localized output distribution) the regularization multiplier can be reduced (less than 1). To get a more spread out distribution, increase the regularization multiplier (greater than 1). Care should be taken if the regularization multiplier is modified to avoid overfitting or underfitting of the model (Phillips et al. 2017).

2.2.2. Testing Maxent Results

The Maxent model provides a robust testing set for measuring uncertainty. Model-based uncertainty methods in Maxent models are found in the form of sensitivity and uncertainty analysis. Confrontational methods include visual tests and statistics-based tests. The Maxent model utilizes both tests in evaluating the outcome of the model. Visual tests can be performed from the layers that are generated from the model that can be rendered as graphics for understanding the model outcomes. Statistics based tests including sensitivity, specificity, threshold dependence plus standard deviation, and regularization help to describe and reduce uncertainty. And finally, Maxent can be run iteratively with different parameters and or constraints to heuristically observe patterns that arise from the model runs.

Indicators of model fitness include the area under the receiver operating characteristics curve (AUC), Omission rate, and True Skill Statistic (TSS). The AUC measures the accuracy of the model in predicting distribution based on sample data. The closer the AUC is to 1, the better the model is at predicting the distribution. In the graphic output provided by Maxent, the mean AUC is shown as the area under the red line and the steeper the angle of increase the closer the AUC value is to 1. An AUC approaching 0.5 means the model cannot predict class separation and therefore cannot predict the distribution at all given the input parameters (the random model is the 1:1 random prediction line depicted in black on the graph in Maxent's output .html file). An AUC approaching 0 indicates reciprocity in the prediction (Narkhede 2018). The receiver operating characteristics (ROC) curve itself is the probability curve measuring the probability the model is a good fit for the data and question being answered. In general, AUC above 70% is considered "sufficiently accurate to be used in conservation planning" (Elith et al. 2006, 141).

Since population size is generally not known but estimated, Maxent cannot produce true occurrence rate per grid cell in the analysis. Sensitivity is a rating of how well the model predicts positive outcomes (or presence). This is the omission rate of the model. Specificity is the measurement of how well the model predicts negative outcomes (or absence). In other words, specificity measures the percentage of absence points that are reported as presence (false negatives) based on modeled probability. This is the commission rate of the model (Phillips 2017; Phillips et al. 2006, Elith et al. 2011; Anderson 2012). However, true commission cannot be measured with presence only data. Sensitivity, however, can be used as a measure of fit along with AUC. Sensitivity and specificity are inversely related. If we decrease the threshold (more positive values) we increase the sensitivity (fewer false negatives) and decrease the specificity (more false positives) (Narkhede 2018). When using AUC as a measure of model performance, it is recommended that omission and commission rates be included in the evaluation where possible (Lobo et al. 2008). Finally, The True Skill Statistic (TSS) is a variation on Kappa that mitigates issues of prevalence that limit the use of Kappa in presence only models like Maxent (Allouche, Tsoar, and Kadmon 2006).

2.2.3. Transferability of Maxent Models

Transferability (also called projection) of SDMs has been an issue of concern in research studies as models built in one geographic space do not always project well to different geographic space and/or time. To maximize transferability, the environmental layers used within the model must align with the requirements of the species but should be broad enough to encompass the entire extent of the originally modeled area and the intended projection area. This alignment is necessary to allow the model to be transferred across space or time for the specific species under review (Anderson 2012; Peterson et al. 2011). Model tuning is recommended to maximize suitability of the model for the species and location being modeled and is especially important when the ability to transfer the resultant model is a desired outcome of the study (Radosavljevic and Anderson 2014).

2.3. Related Research

Background research for this thesis included review of research in the following areas: Studies which similarly used the Maxent model to predict invasive species distribution, research on model transferability, research on modeling distribution of cogongrass specifically, and research on Maxent model parameters. There have been several studies of invasive species, including cogongrass, utilizing the Maxent model that are useful background references for this study.

Amanda West et al. (2016) endeavored to predict invasive species distribution of cheatgrass (*Bromus tectorum* L.) utilizing Maxent. West states that presence-only models are rarely evaluated against real field data, therefore, the authors determined to test field data collected over a period of time against the Maxent presence only model. Presence data collected between 2007 and 2013 were used as the sample data for the study. West et al. ran a Maxent model across the area in 2007 using limited real data. Then, the Maxent model was rerun using the new data from 2008 to 2013 using same parameters as used in 2007 to test the accuracy of the previous results. A new model with updated parameters was also run and was tested with the same sample dataset collected between 2008 and 2013. West used area under the curve (AUC), percent correctly classified (PCC), sensitivity, specificity, and true skill statistic (TSS) to evaluate and validate the models. The West et al. study concludes that the Maxent model is a good fit for measuring the distribution of invasive cheatgrass in the Rocky Mountain National Park .

Ervin and Holly (2011) performed a similar study at Mississippi State University on cogongrass in southeastern Mississippi in an attempt to determine if the Maxent model they designed for the Mississippi varietal would transfer to appropriately predict the distribution of the Alabama varietal testing their Mississippi model against three subsets of Alabama cogongrass data from the same geographic area but collected in three different years. They determined that there was low transferability of the Mississippi model from Mississippi to Alabama but noted several potential reasons for this low transferability including the landscapes that were focused on (Mississippi focused on roadways while Alabama focused on managed timberland) and discrepancies in soils. These concerns related to environmental factors affecting transferability support the use of landcover data, distance to roads layer data and soils related variables in the Maxent model generated in the current study. Another important point here is that the lineage of cogongrass in Mississippi is from the Philippines while the lineage of Alabama cogongrass is from Japan. That genetic difference may also be a factor in how the species responds to environmental factors within the study (Lucardi, Wallace, and Ervin 2011). A study of genetic impact is out of scope for this analysis; however, it is worth noting as lineage can play a role in species response to environmental stimuli. Importantly, Ervin and Holly's Mississippi dataset was collected in a different fashion and for a different purpose (different landscape focus as mentioned above) than the Alabama dataset used in that study. This could have had an impact on the poor transferability of the model.

A 2005 case study on cogongrass published by the US. Forest Service indicated that cogongrass may outcompete native species in poor soils due to its dense rhizome mat (Howard 2005) allowing the invasive to restrict access to soil nutrients and water for native grasses (Howard 2005; Lippincott 1997). Cogongrass rhizomes have been shown to be present in the top 15cm of fine textured soils or top 40cm of coarse textured soils (MacDonald 2004). Howard also noted that native species that outcompete cogongrass successfully generally have deeper root systems or taller crowns, although MacDonald noted that cogongrass rhizomes formation may be present to depths of 120cm (MacDonald 2004, Holm 1977). This study supports the use of soils related variables such as depth to soil restrictive layer, soil texture layers, and drainage class as environmental covariates for use with Maxent.

A 2000 study by King and Grace examined soil moisture content's effect on cogongrass seedling germination and growth, testing soil saturation ranging from dry to inundated.

Measurements of plant height and number of shoots were used to define seedling growth rate and germination success respectively. This study found that cogongrass seedling germination was weakest (reduced by 74%) when soils were inundated and that growth became increasingly restricted, especially for smaller seedlings, as soil saturation increased. The authors suggested, based on the results of their study that soil inundation in the early stages of cogongrass establishment could restrict invasion by seed.

Roads as a pathway for seed dispersal was reviewed in a 2017 study by Rauschert, Mortensen and Bloser. In this study, the authors followed physical seed dispersal of *Carthamus tinctorius* L.(safflower) seeds, by routine rural road maintenance equipment, specifically by road graders, on rural dirt roads. Safflower seeds were used in this study as the use of invasive seed was restricted. The authors placed four patches of 5000 painted seeds in a grid across a rural road that was planned to be graded using a typical three pass approach. Then, immediately following the road grading event, Rauschert, Mortensen and Bloser measured the distance that seeds traveled based on seed starting location and ending location. The study found that 41.8% of seeds moved between 10 and 50 meters and only 1.6% traveled greater than 50 meters with a maximum movement of 273m. This study focused on the physical movement of seed by road maintenance equipment, however hitchhiking of seeds on vehicles and wind dispersal along roadways was not included in this study but has been identified as additional key pathway for dispersal related to transportation corridors. This study as well as mention of roads ad vectors for dispersal in other studies provides incentive to include distance to road as an environmental variable within the current study on cogongrass (Rauschert, Mortensen and Bloser 2017).

A study on woody shrubs as a barrier to wind dispersal of cogongrass seed (Yager, Miller, and Jones 2011) was performed at Camp Shelby Training Site in Mississippi. This study tested the travel distance of cogongrass spikelets (seed removed) released along three sites which each contained blocks of pine-shrub forest and pine-tallgrass forest. The goal of the study was to determine if forests with a woody shrub mid-story reduced the dispersal of cogongrass spikelets and therefore reduced invasive introduction to forested areas along roadways. The study found that although mean dispersal distance of cogongrass spikelets was not significantly different between the two forest types, that more spikelets traveled further in the pine-tallgrass forest (25% dispersed further than 5m) than did in the pine-shrub forest (8% dispersed further than 5m) and the mean maximum dispersal distance was greater in the pine-tallgrass forest (37m) than in the pine-shrub forest (23m). The study concluded that cogongrass dense woody shrub vegetation along forest edges may impede the wind dispersal of cogongrass spikelets and subsequent invasion of the species to the forest interior however it does point out that, in areas where cogongrass is already present, infestation growth may still occur via vegetative spread as the species some shows tolerance to shade.

A 2018 study reviewed the impact of grain size of predictor variables on the accuracy and transferability of SDM models specifically using Maxent to test transferability for an invasive plant species (*Rhododendron ponticum* (L.)) in Wales, U.K. (Manzoor, Griffiths, and Lukac 2018). The authors noted that the selection of grain size in SDMs is often dependent on the availability of appropriate predictor variable (environmental covariate) data and the available resolution of that data. As noted by the authors, finer grain size allows for more detailed and potentially more accurate prediction of suitable environmental habitat and courser grain size inhibits habitat delineation. Maxent requires all environmental covariates to utilize the same grain size and therefore, The authors focused this study on comparing Maxent outputs of three models. The modeled grain sizes were 1km, 300m, and 50m. For the 50m model, biophysical variables of Altitude, Aspect, Slope, Land Cover, and Distance from water channels were used in the model as environmental covariates. These same datasets were resampled to 300m to be used in the 300m model and also resampled to 1km for the 1km model (see Manzoor, Griffiths, and Lukac 2018 for details on methods used). For the 1km model, bioclimate variables were also included as is common in SDM studies (Manzoor, Griffiths, and Lukac 2018). Model transferability at all three grain sizes was also tested to determine if grain size has an impact on the transferability of the model.

This study used the Continuous Boyce Index (CBI) to test transferability. The study results show that CBI improves as grainsize is reduced in both the training model area as well as the transfer test area. In the training model area, the CBI improved from 0.825 for the 1km model to 0.895 for the 300m model to 0.964 for the 50m model. In the transfer test area, the CBI improved from 0.65 for the 1km model to 0.90 for the 300m model but dropped to 0.77 for the 50m model. The reduction in CBI between the 300m and 50m models in the transfer test area was attributed to differences in range and topography of the two geographic areas of study. The authors concluded that, although the use of climate data is widely used in SDMs and in many cases this is justified, biophysical variables based on the biology and ecology of the species being studied as well as the spatial extent of the study area may be more important for localized studies. Therefore, the authors suggest that the use of course grained climate datasets should be considered in reference to their overall importance to the specific species and geographic extent of the study, and that the inclusion of these climate datasets can produce less accurate SDMs due to the required coarser grain size of the model.

Chapter 3 Data and Methods

As discussed in Chapter 1, this study focuses on *Imperata cylindrica* (L.) Beauv., more commonly known as cogongrass, which is a highly invasive species with high tolerance to a broad range of environmental conditions. Maxent was utilized to model the predicted potential distribution of cogongrass infestation given suitable conditions for the AFC's Work Unit 11 (Model Study Area) and then transferred to Work Units 12 (Test Area 1) and 8 (Test Area 2) to test model transferability across the state.

As described in Figure 7, the use of Maxent for species distribution modeling has four key steps. First, the species was researched to ensure that environmental variables selected for the model are relevant to the species and study location. This is discussed in more detail in the Data Description section below. Second, the datasets were cleansed, and data layers were prepared. This included both the species presence data as well as the environmental variables. Third, the Maxent model was run at default and then tuning occurred to ensure parameter settings were appropriate for the study. Finally, the Maxent results were evaluated and model performance was determined.

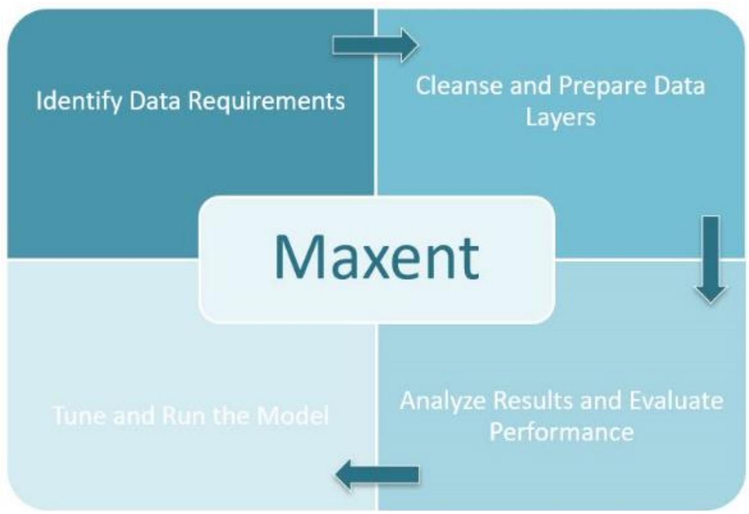


Figure 7: High level overview of Maxent steps

The ultimate goal of this study was to determine if Maxent is an appropriate tool to predict cogongrass distribution and, if so, to determine if a locally constructed model could be transferred to other areas within the state successfully. This study endeavors to create a model that can be transferred to each work unit and reliably predict cogongrass infestation locations to help guide survey and eradication efforts by the AFC.

3.1. Study Area

The primary Model Study Area (Figure 8) is the Alabama Forestry Commission's (AFC) Work Unit 11 encompassing 10,902 km^2 with an average point per km^2 of 0.85. This study area includes Choctaw, Marengo, Clarke, and Washington counties. The Model Study Area was chosen because the area contains a large verified point location dataset to use in the model (9,242 points) and this AFC Work Unit contains the transferability study area from the Ervin and Holly 2011 study (Clarke County, Alabama) that initially sparked my interest in model transferability. Transferability of the resultant model was then tested against similar as well as dissimilar Work Units within the state. Test Area 1 (Figure 9) consists of AFC Work Unit 12 which encompasses 7,306 km^2 with 6826 presence points that fall within the boundary of the study area and an average point per km^2 of 0.93. Test Area 2 (Figure 10) consists of AFC Work Unit 8 which encompasses 8,088 km^2 with only 78 presence points that fall within the boundary of the study area and an average point per km^2 of 0.01. Table 3 shows the comparison of area and number of points in each of the Alabama Work Units.

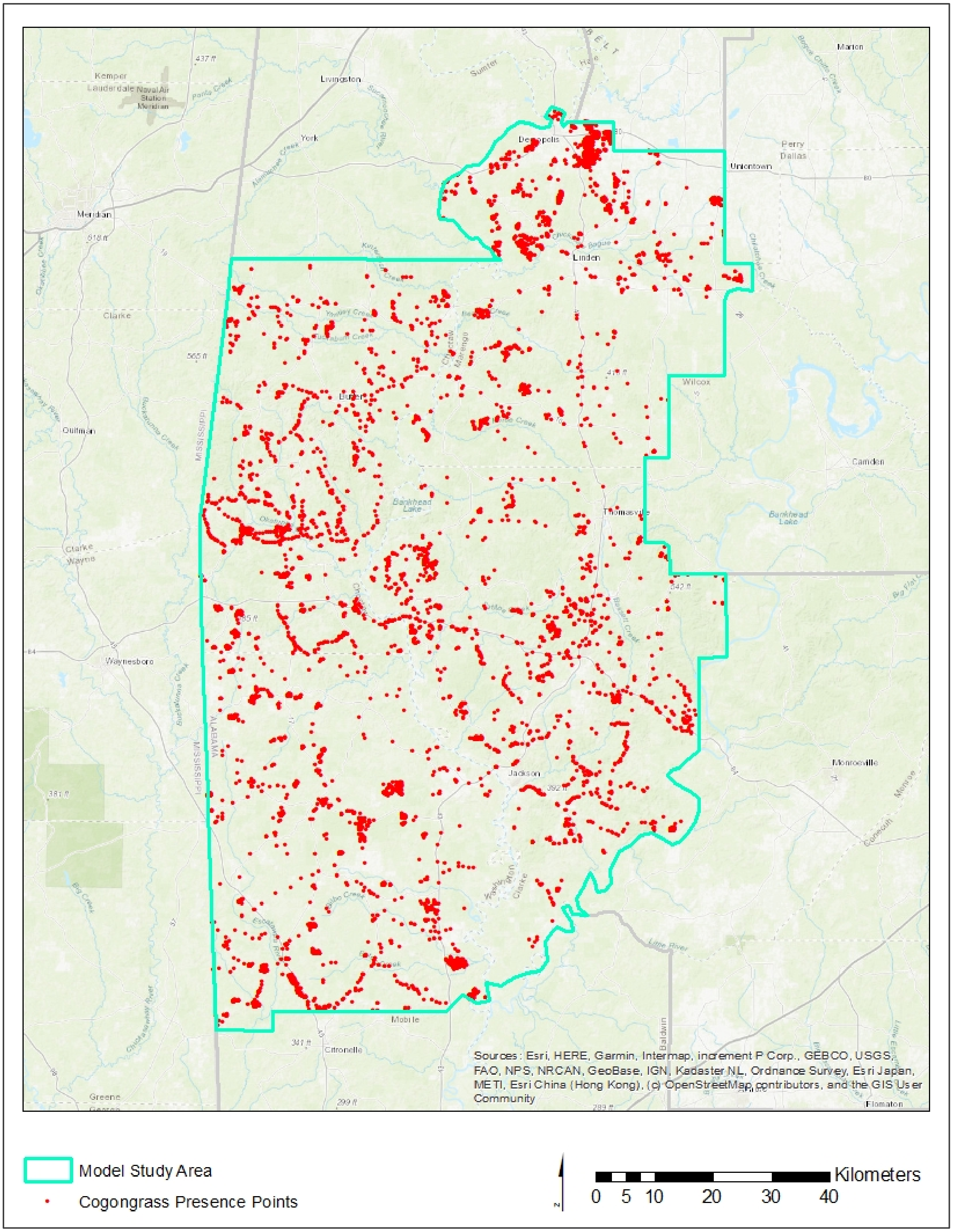


Figure 8: Model Study Area with Cogongrass Infestation Presence Point Locations Identified

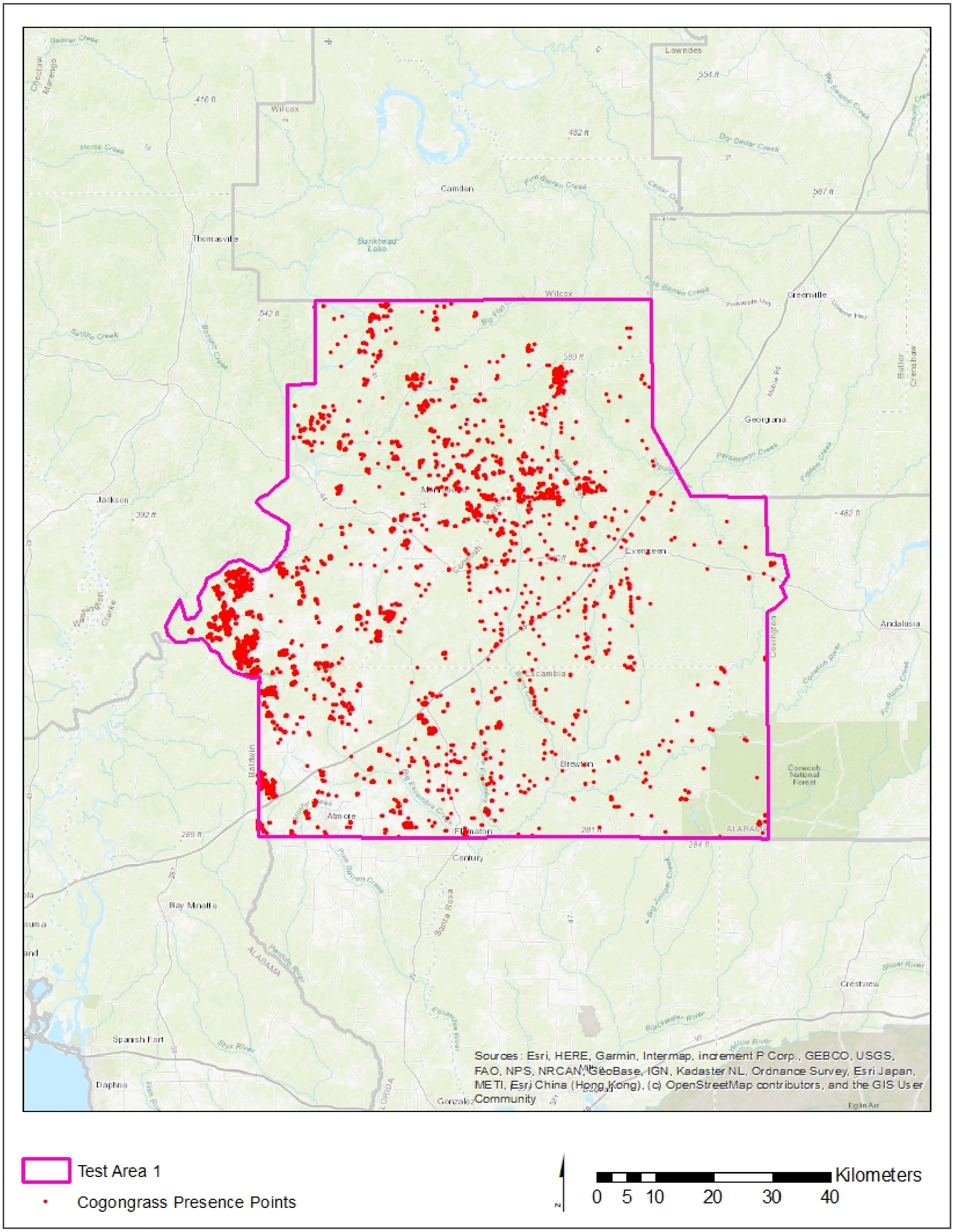


Figure 9: Test Area 1 with Cogongrass Infestation Presence Point Locations Identified

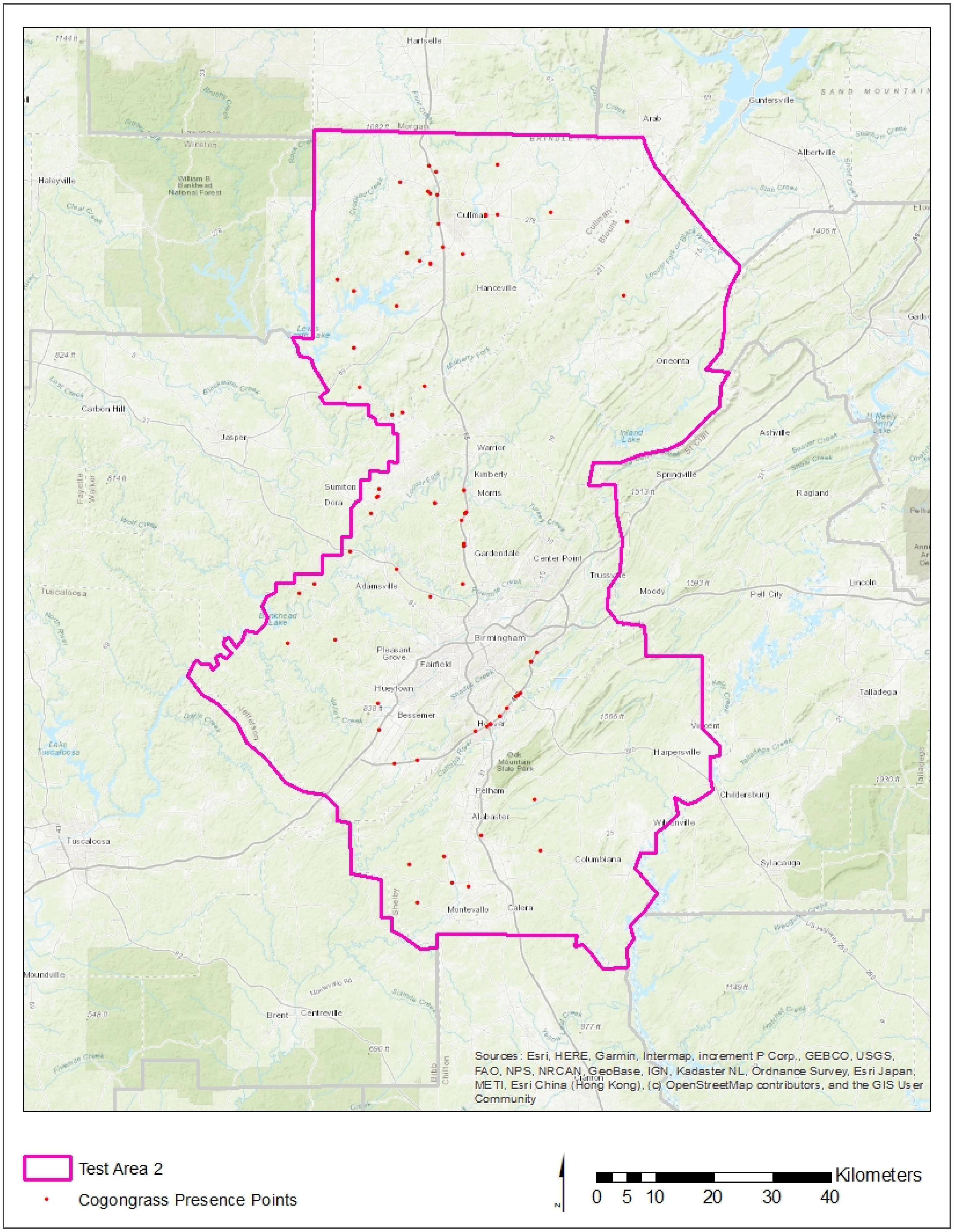


Figure 10: Test Area 2 with Cogongrass Infestation Presence Point Locations Identified

Table 3: Count of verified point locations, total km^2, and average points per km^2 in each AFC Work Unit.

Work Unit	Count of Points	Square Kilometers	Average Points per Square Kilometer
1	5	6,468	0.00
2	47	6,688	0.01
3	2,355	10,378	0.23
4	4,253	7,867	0.54
5	9	7,009	0.00
6	22	5,840	0.00
7	6	8,347	0.00
8	78	8,088	0.01
9	13	6,689	0.00
10	1,898	6,700	0.28
11	9,242	10,902	0.85
12	6,826	7,306	0.93
13	7,389	7,302	1.01
14	256	7,219	0.04
15	435	5,330	0.08
16	88	6,425	0.01
17	1,557	5,916	0.26
18	262	6,702	0.04
Total	**34,741**	**131,176**	
Average	**1,930**	**7,288**	**0.24**

As can be seen in Table 4, Work Unit 11 is mostly rural with 50% upland forest/woodlands and 27% floodplain forest. Eight percent of this area is in agricultural use and only three percent is developed. Like the Model Study Area, Test Area 1 is very rural in nature. This area includes 47% upland forest/woodlands, 22% floodplain forest, 15% agricultural use, and is three percent developed. Test Area 1 has the closest number of presence points per km^2 to the Model Study Area. Based on visual inspection of the environmental layers using ArcGIS 10.6, it was determined that Test Area 1 has similar distribution of land use, PctClay, PctSilt, PctSand, drainage class and PctCanopy with higher pH and slightly more agricultural use. (See Appendix A and Appendix B for maps of each environmental layer used in the analysis). It was therefore expected that Test Area 1 would display a similar predicted distribution to the Model Study Area.

Table 4: Ecological System categories for comparison of study areas' land use differences.

Category	Model Study Area % of area	Test Area 1 % of area	Test Area 2 % of area
Forest/Woodlands	49.99	46.88	52.34
Floodplain Forest	26.54	22.24	0.14
Agriculture	8.07	14.64	20.84
Developed	3.18	3.09	15.05
Disturbed	10.99	12.39	8.19
Water	1.18	0.68	2.80
Other	0.06	0.07	0.64

Test Area 2 includes the city of Birmingham, the largest city in the state, and is the most urban area within the state of Alabama. This area's land use includes: 52% upland forest/woodlands, <1% floodplain forest, 21% in agriculture and is 15% developed (Table 4). Again, based on visual inspection of the environmental layers, it was determined that Test Area 2 has significantly greater variability in depth to restrictive layer, lower PctCanopy, more PctClay and PctSilt and less PctSand than the Model Study Area. Test Area 2 was included to test transferability of the model defined for the Model Study Area to an area of dissimilar environmental makeup.

3.2. Scale of Study

A key component of gridded data, such as the ASCII files used by Maxent, is the grain size which represents the spatial resolution of the layers to be included in the analysis (Manzoor, Griffiths, and Lukac 2018). For a model such as Maxent to function properly, all layers must be set to the same spatial resolution. The resolution selected for this study was 30m resolution which is the resolution of the primary datasets used to build and test the model. The Soils and Land Use datasets are both native 30m resolution.

I considered testing at 200m and 800m to allow for the inclusion of Parameter-elevation Regressions on Independent Slopes Model (PRISM) climate datasets. According to the PRISM Climate group website at Oregon State University, PRISM data is provided by the PRISM Climate Group which gathers climate data from monitoring networks and develops spatial climate datasets to be used to show short- and long-term climate patterns. Although the use of climate data in species distribution modeling is common (Manzoor, Griffiths, and Lukac 2018), it was determined to be unnecessary in this study as the scale of data available was too coarse to provide adequate detail to inform the model. Also, given that cogongrass is highly tolerant to a wide range of climatic conditions, and the climate of the state meets this range for all pertinent climate data in all but the extreme northeastern portion of the state, inclusion of climate data was determined to be superfluous.

3.3. Data Description

A wide range of environmental variables are available both publicly and privately for use in SDMs such as Maxent. To minimize potential overfitting of the model, care was taken to select only variables that were relevant to the species and study location and to reduce redundancy in variables where possible. For greater model relevance and to minimize correlation between variables used, it is advised to select environmental variables that are relevant to the species being studied (Manzoor, Griffiths, and Lukac 2018; Radosavljevic and Anderson 2014).

When employing data from multiple sources, however, ensuring proper alignment of the data can be difficult. Maxent requires that all environmental layers used in a given model match in geographic extent, grid cell size, and projection (Elith et al. 2011; Phillips 2017; Ervin and

Holly 2011). To this end, all environmental variables used within this study were sampled at a 30m by 30m grid cell size in the North American Datum (NAD) 1983 UTM Zone 16N projection. The species point location dataset was also projected to NAD 1983 UTM Zone 16N to match the projection of the environmental variables.

Datasets to be used within this analysis include the verified point location dataset of cogongrass from the Alabama Forestry Commission, USGS GAP Land Cover data set, USDA Soils data, and four roads datasets provided by Silvics Solutions LLC comprising Local, State, US Highways, and Interstate features (see Table 5). For reference and visualization purposes State, County, and Bing Maps base maps were also used in the study.

Table 5: Datasets used in this study

Dataset	Source	Description
Cogongrass verified infestation point location dataset	Alabama Forestry Commission	Point location dataset for all verified infestation locations in the state of Alabama as reviewed and verified by the Alabama Forestry Commission. The publication of this dataset is 4/19/2018. This dataset can be acquired through direct request from the Alabama Forestry Commission.
Distance to Nearest Road Feature	Silvics Solutions LLC	Calculated using the Euclidean Distance tool in ArcGIS 10.6 from four road layers provided by Silvics Solutions LLC.
USGS GAP Land Cover Data Set	Databasin.org (https://databasin.org/datasets/e6c2c82715be44bba3579fa6010acfd5)	"The USGS GAP Land Cover Data Set includes detailed vegetation and land use patterns for the continental United States. The data set incorporates the Ecological System classification system developed by NatureServe to represent natural and semi-natural land cover." (USGS website). Projection = NAD_1983_Albers.
USDA Soils data	United States Department of Agriculture Soil Survey Geographic Database (SSURGO)	Soils data as collected by the National Cooperative Soil Survey. The survey is broken down into map units (polygons) describing the soils and other components of the soils such as productivity, and soil horizons. The information was collected at scales ranging from 1:12,000 to 1:63,360. Projection = World Geodetic System 1984 in units of decimal degrees.

All datasets were projected to NAD 83 UTM Zone 16 North and clipped to the boundary of the state of Alabama prior to the outset of the study. Any data falling outside of that boundary was removed from this analysis. All environmental variable (covariate) datasets are publicly available for download with the exception of the specific road layers used for the Euclidean Distance calculation, however similar roads datasets are available publicly. See Table 6 for the specific source of each environmental layer used in the study. Maps showing each of these layers are included in Appendix A and B.

Table 6: Environmental variables used in the study along with their layer name abbreviation and specific source and tool used for creation where applicable

Variable	Abbreviation	Source
Percent Canopy Cover	PctCanopy	nlcd_2011_USFS_tree_canopy_2011_edition_2016_02_08_cartographic
Ecological System	EcolSys	GAP Land Cover Data for Alabama, USA (gap_30m_al)
Distance to Roads	Distance	Roads layer provided by Silvics Solutions, LLC. Distance calculated using the Euclidean Distance tool in the Spatial Analyst toolbox in ArcGIS 10.6
Soil pH	pH	gSSURGO Soils Data Development Tools toolbox. gSSURGO Mapping Toolset. Create Soil Map tool for ArcGIS
Soil Particle Size	PartSize	MUPOLYGON layer from gSSURGO_g_al database
Drainage Class	DC	gSSURGO Soils Data Development Tools toolbox. gSSURGO Mapping Toolset. Create Soil Map tool for ArcGIS
Depth to Restrictive Layer	Bed	gSSURGO Soils Data Development Tools toolbox. gSSURGO Mapping Toolset. Create Soil Map tool for ArcGIS
Percent Clay Content	PctClay	gSSURGO Soils Data Development Tools toolbox. gSSURGO Mapping Toolset. Create Soil Map tool for ArcGIS
Percent Silt Content	PctSilt	gSSURGO Soils Data Development Tools toolbox. gSSURGO Mapping Toolset. Create Soil Map tool for ArcGIS
Percent Sand Content	PctSand	gSSURGO Soils Data Development Tools toolbox. gSSURGO Mapping Toolset. Create Soil Map tool for ArcGIS

3.3.1. Species Presence Data

The species presence data used in this study consisted of the verified point location dataset for Cogongrass (*Imperata cylindrica* (L.) Beauv.) as provided by the Alabama Forestry Commission. Specifically, this data was provided by Dana Stone, Forest Health Coordinator, Alabama Forestry Commission, Montgomery, AL and was provided in shapefile format. Figure 11 shows the entire set of points. This is a very robust dataset including over fifty-four thousand reported points and 34,771 field verified points.

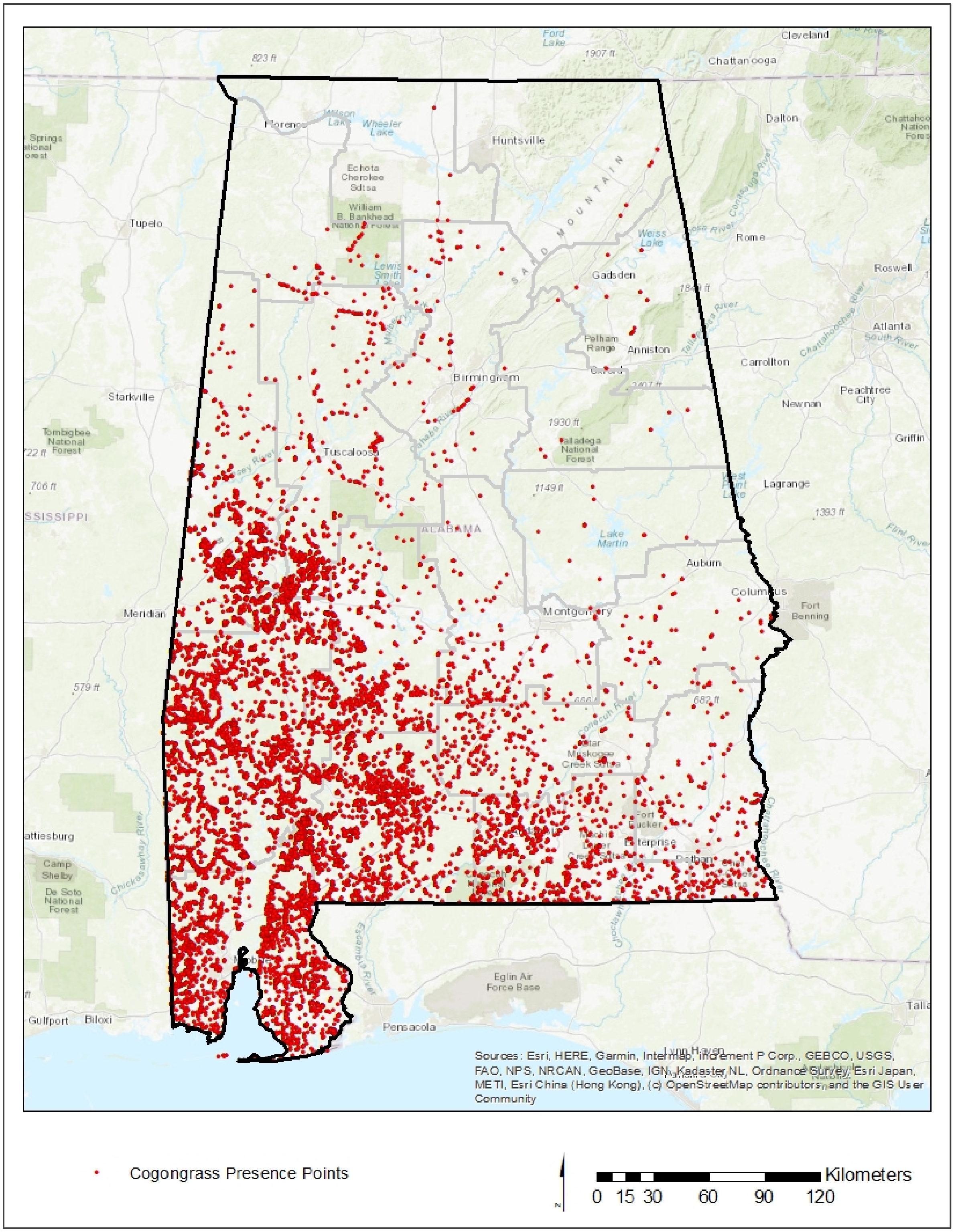


Figure 11: Alabama Forestry Commission field verified cogongrass infestation locations in the state of Alabama

The presence point dataset was cleansed prior to use in this study. First the dataset was reduced to only those points that have been field verified, the dataset was then projected to NAD 1983 UTM Zone 16N and the Extract by Mask tool was used to remove all presence points that fell outside of the State of Alabama. This tool was used extensively in the data preparation phase of this study and therefore warrants a brief explanation of its function.

The Extract by Mask tool in the Spatial Analyst Toolset is used to extract cells from an input raster that correspond to the area defined by a mask layer which can be vector or raster. This tool is used to ensure that the output raster has the exact same cell count in number of columns and rows as the mask layer. It is a requirement of Maxent that all environmental layers have the same header values in the ASCII files used in the model run. In Maxent, if layers do not have the same values in the header, the model will error out and cannot be run until the extent of each ACSII file matches exactly. Since the mask feature specified was a vector layer rather than a raster layer, the tool internally converts the vector to a raster and marks any point cell whose cell center point falls outside of the original vector boundary as No Data (Esri 2019).

The points were then further extracted to create separate environmental data layers for the Model Study Area and the transferability test areas. Once the final point location datasets needed for the model were generated, new X and Y coordinate columns were added to the attribute table and the X and Y values were calculated in meters. This table was then exported to Microsoft Excel using the Table to Excel tool and the resultant Excel file was converted to a comma-separated values (CSV) file format for use in Maxent. Steps for preparation of the presence point location dataset for use in Maxent are outlined in Appendix C.

3.3.2. Soils Data Overview

Cogongrass' tolerance to a wide range of environmental conditions (Terry et al. 1997; Howard 2005) makes selection of appropriate environmental layers for analysis tricky. The ultimate goal of the study is to provide a model that is transferable to areas across the state of Alabama and therefore environmental variables considered for the study need to be both granular enough to contribute usable outcomes and broad enough to be applicable across the entire state. Descriptions of the soils related environmental variables included in this study can be found in Table 7 and details on each one are given in the following paragraphs.

Table 7: Soil variables used in the study

Soils Variable	Description
Bed	Depth in centimeters to the layer that impedes water and air movement or restricts root growth within the soil (depth to restrictive layer)
DC	Drainage class is an indication of the soil's wetness and/or saturation
PartSize	Particle size is the general classification of the soil texture as determined by grain size for the topmost horizon of soil (standards used by the U.S. Department of Agriculture). Terms defined according to % of sand, silt and clay.
PctClay	Percent clay is the percentage by weight of soil with mineral particles less than 0.002 mm in diameter.
PctSand	Percent sand is the percentage by weight of soil with mineral particles ranging from 0.05mm to 2mm in diameter
PctSilt	Percent silt is the percentage by weight of soil with mineral that range from 0.002mm to 0.05mm in diameter
pH	Soil pH is a measure of the acidity or alkalinity of the soil using the 1:1 water method of measurement.

Depth to Restrictive Layer (Bed) quantifies the depth in centimeters to the layer that impedes water and air movement or restricts root growth within the soil. According to the metadata layer properties associated with the gSSURGO_CreateSoilMap.py script used to generate this layer, the restrictive layer is a continuous layer and can be a physical, chemical, or thermal barrier. The fire case study discussed in Section 2.3 indicated that cogongrass may outcompete native species in poor soils due to its dense rhizome mat (Howard 2005) allowing the invasive to restrict access to soil nutrients and water for native grasses (Howard 2005; Lippincott 1997). Depth to soil restrictive layer as well as drainage class were selected as environmental variables within this study as proxy for these considerations. Figure 12 shows thumbnail images of this layer. The values range from greater than 201 cm (bright green) to 0 cm (red). Note how Test Area 2 is very different from the other two areas as it has much shallower depth to restrictive layer in much of its geographic area. For this and all subsequent thumbnail images in this section, large images of these layers are included in Appendix A: Soils Related Environmental Covariate Maps.

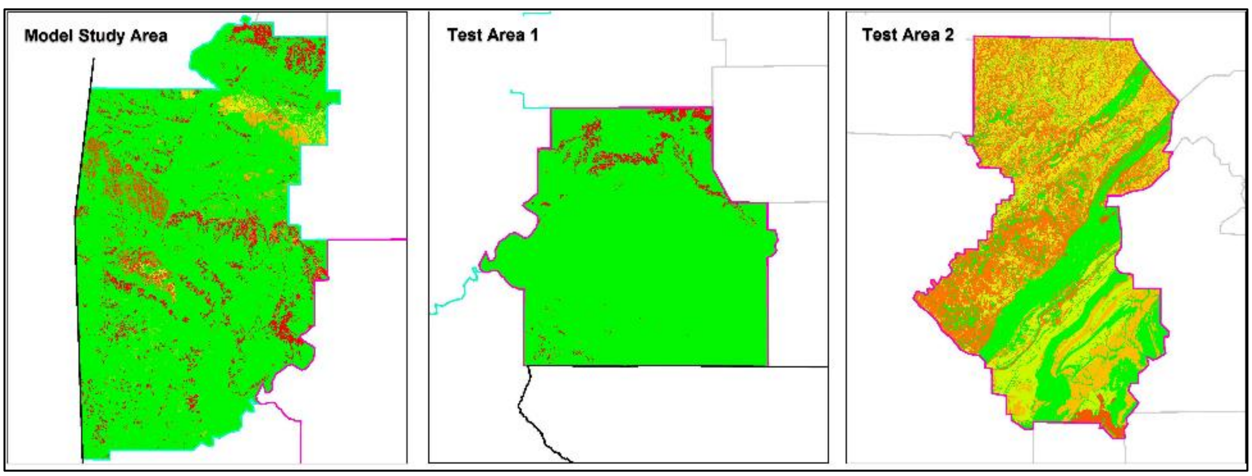


Figure 12: Depth to restrictive layer thumbnail images for each of the study areas. See Appendix A for larger images.

Drainage Class (DC) is a representation of moisture content in the soil in its natural condition. There are seven subclasses which range from excessively drained to very poorly drained within the drainage class variable (www.epa.gov/enviroatlas). A study by King and Grace showed that high water levels restricted cogongrass seedling growth and the seedlings germination was reduced by 74% when soils were flooded (King and Grace 2000). The Model Study Area had poorly drained soils over 26% of its total area, whereas Test Areas 1 and 2 had

15% and 4% of their areas consisting of poorly drained soils respectively. In contrast, the Model Study Area had 68% of the total area covered with well drained soils, whereas Test Areas 1 and 2 had 73% and 86% of total area consisting of well drained soils respectively. Figure 13 shows thumbnail images of this layer. Well-drained soils are shown in mossy greens and poorly drained soils are shown in blues. Large images of these layers are included in Appendix A: Soils Related Environmental Covariate Maps.

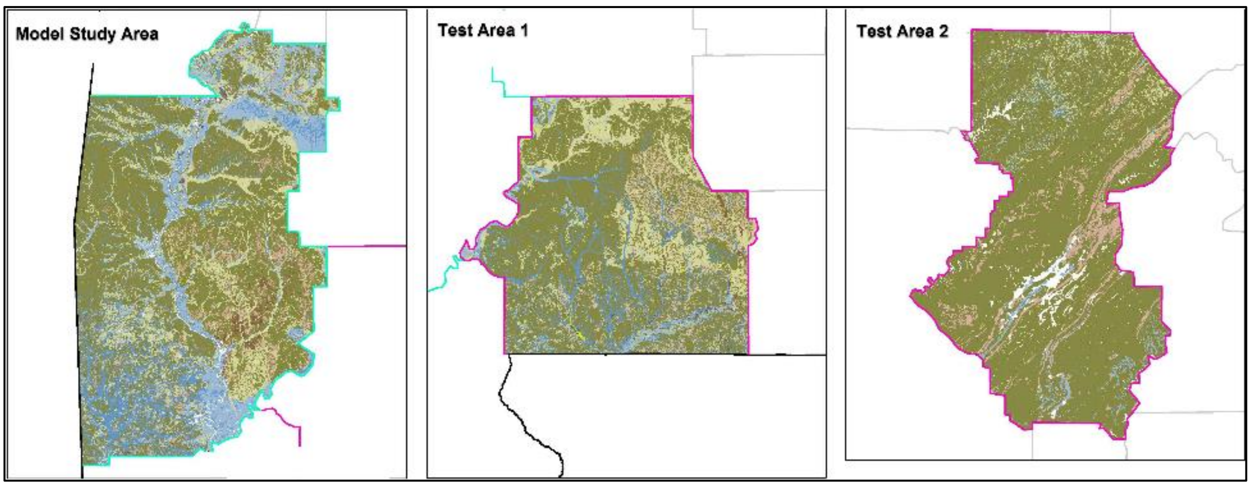


Figure 13: Drainage class thumbnail images for each of the three study areas. See Appendix A for larger images

Particle size (PartSize) represents a general classification (grouping) of the soil texture as determined by grain size for the topmost horizon of soil using the standards used by the U.S. Department of Agriculture. This grouping places soils with somewhat similar properties in the same particle size class and is helpful when a general view of soil texture is needed. PartSize classification is defined according to percent of sand, silt, and clay as shown on Figure 14, therefore there is some correlation between this variable and the individual percentages for sand, silt and clay used within the model. Soil particle size, both in broad categorical terms as well as percent clay, silt, and sand, were included in this analysis.

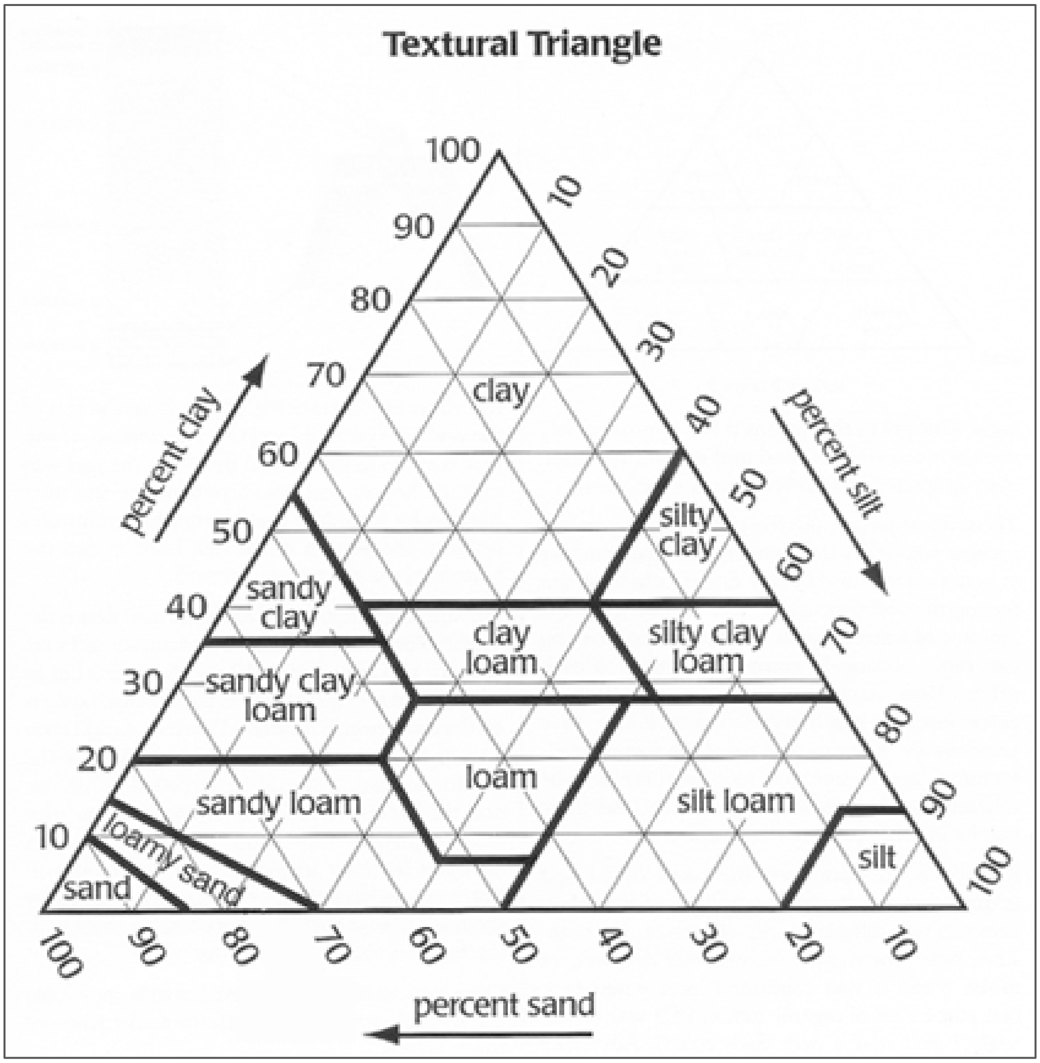


Figure 14: The soil texture triangle is used to convert the relative amounts of clay, silt, and sand in the soil into texture classes. For example, a soil that is 20% clay, 30% silt, and 50% sand is a Silty Loam. Image courtesy of Grow it Organically website, https://www.grow-it-organically.com/facts-about-soil.html

Figure 15 shows thumbnail images of soil particle size . Particle size coloration in the maps indicate PartSize categories as clayey soils (oranges), Silty soils (blues), Loams (greens) and sandy soils (browns). Note that Test Area 2 has much more fine-loamy soil (light green) than the Model Study Area or Test Area 1. Large images of these layers are included in Appendix A: Soils Related Environmental Covariate Maps.

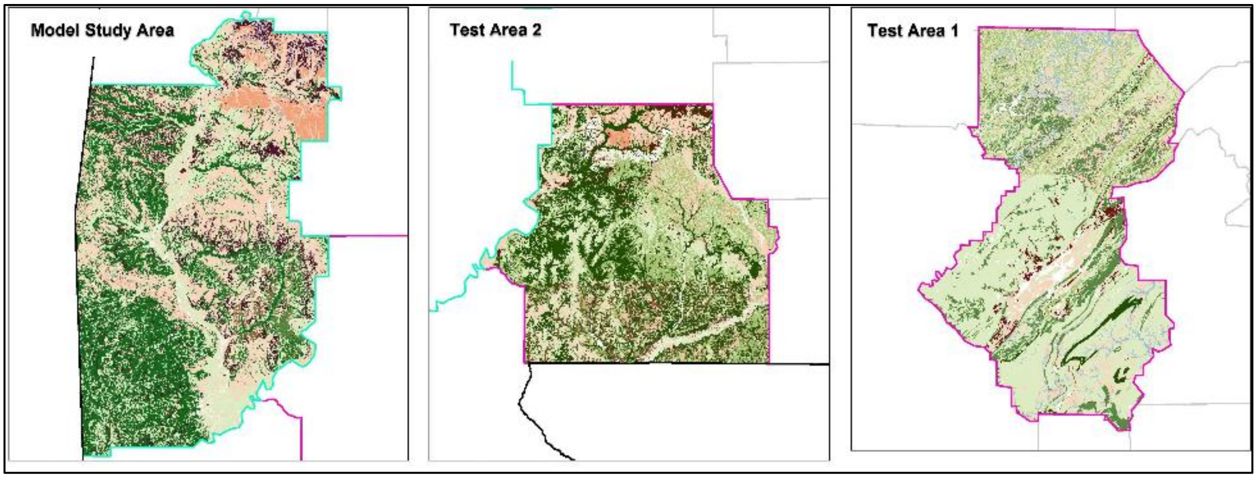


Figure 15: Particle size thumbnail images for each of the study areas. See Appendix A for larger images.

The Percent Clay Content (PctClay) variable represents the percentage by weight of soil with mineral particles less than 0.002 mm in diameter. The percentage and kind of clay found in soil has significant impact on land use, drainage, fertility, etc. Figure 16 shows thumbnail images of this layer. PctClay ranges across the study areas from 0% to 71.6% clay content where darker color indicates higher percentages. Large images of these layers are included in Appendix A: Soils Related Environmental Covariate Maps.

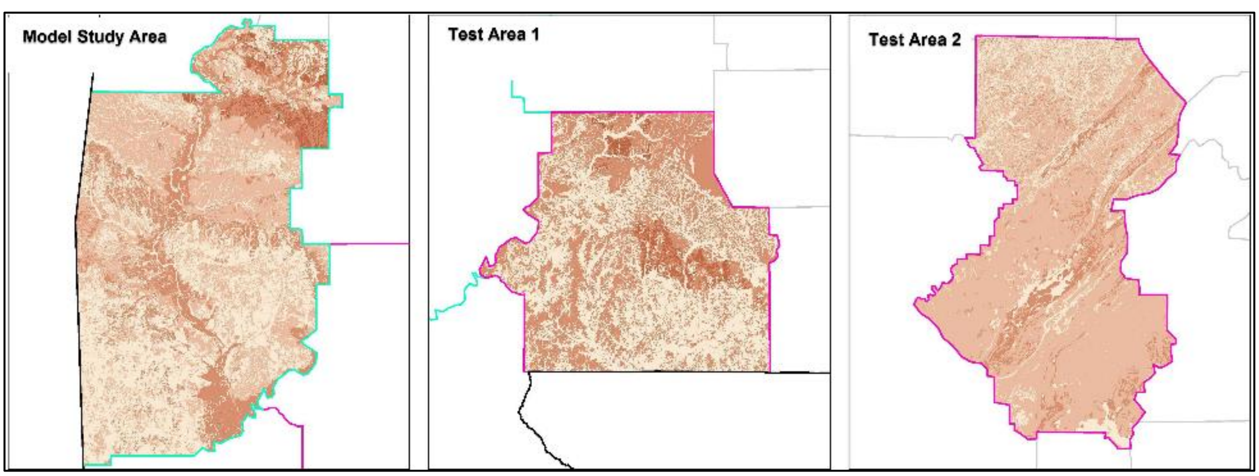


Figure 16: Percent clay content thumbnail images for each of the study areas. See Appendix A for larger images.

The Percent Sand Content (PctSand) variable represents the percentage by weight of soil with mineral particles ranging from 0.05mm to 2mm in diameter. Figure 17 shows thumbnail images of this layer. PctSand ranges across the study areas from 0% to 94.1% sand content where darker color indicates higher percentage. Large images of these layers are included in Appendix A: Soils Related Environmental Covariate Maps.

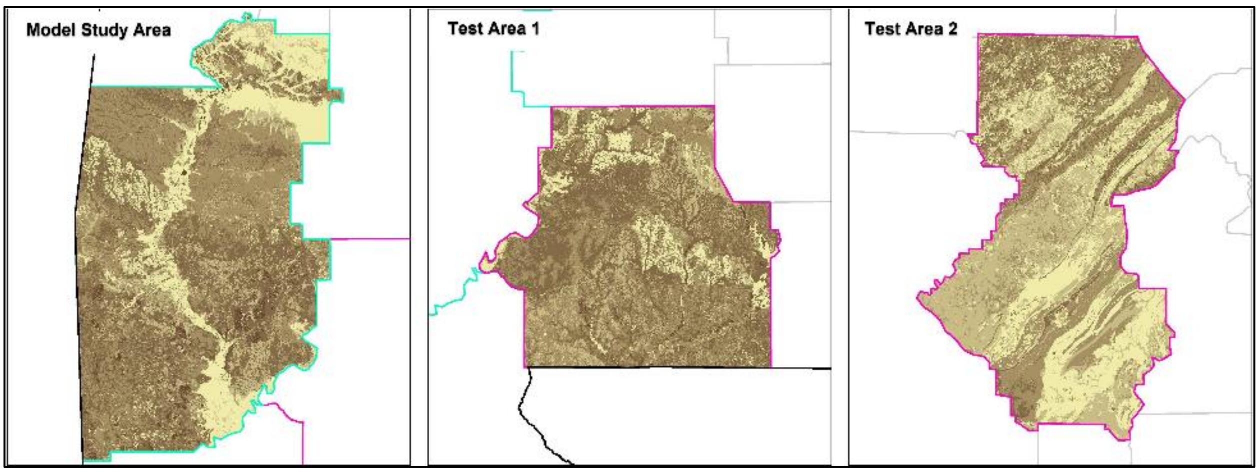


Figure 17: Percent sand content thumbnail images for each of the study areas. Appendix A for larger images

The Percent Silt Content (PctSilt) variable represents the percent of mineral soil particles that range from 0.002mm to 0.05mm in diameter. Figure 18 shows thumbnail images of this layer. PctSilt ranges across the study areas from 0% to 66% silt content where darker color indicates higher percentage. Large images of these layers are included in Appendix A: Soils Related Environmental Covariate Maps.

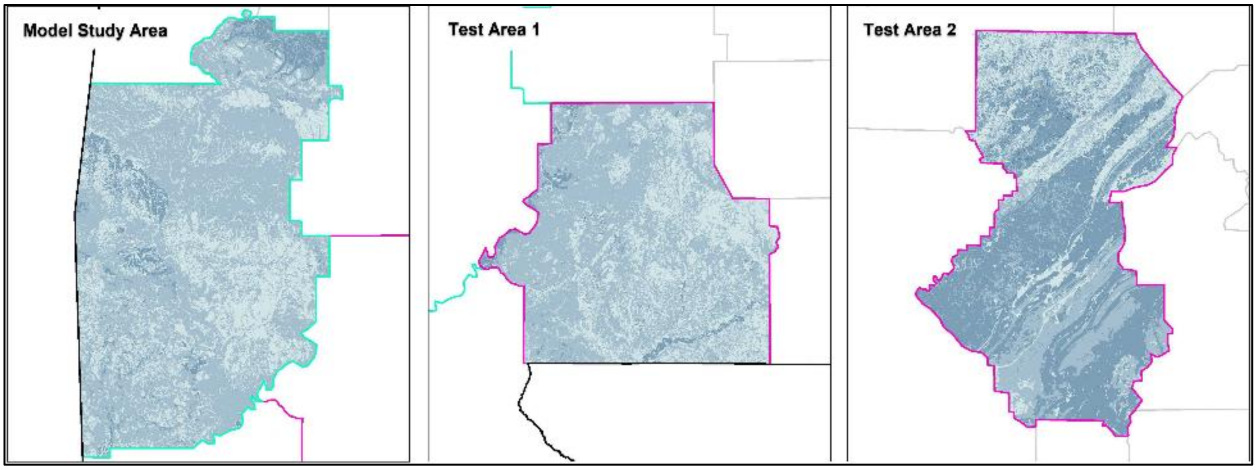


Figure 18: Percent silt content thumbnail images for each of the study areas. Appendix A for larger images.

The soil pH (pH) variable represents a measure of the acidity or alkalinity of the soil using the 1:1 water method of measurement. Cogongrass has been shown to grow best in relatively acidic soils (pH of 4.7) and a study by Wilcut et al. (1988a) states that cogongrass grew better in soils of pH 4.7 than at pH 6.7. In that study, soil pH of 6.7 was chosen to represent typical soil pH of cultivated fields. Seed germination has also been shown to increase at pH less than 5.0 (Sajise 1976). Although cogongrass has stronger growth rates in more acidic soils, the species can grow in a broad range of pH values at sub-optimal growth rates.

The soil layers included in this study were selected for their relevance to the biological and ecological niche of cogongrass. Several studies have raised the importance of soil pH, not necessarily on the presence of cogongrass, but on the health of the species in its environment (Ervin and Holly 2011; MacDonald 2004; Eussen and Wirjahardja 1973). Ervin and Holly suggested that soil pH may play a larger role in cogongrass infestation in their transferability test site in Clarke county Alabama (Ervin and Holly 2011), therefore it was determined that soil pH would be a useful environmental variable to include in this study.

Figure 19 shows thumbnail images of this layer. Soil pH values range across the study areas from 0 to 8.3 where red is the most acidic and blue is the most alkaline. The full pH scale ranges from 0 to 10. Large images of these layers are included in Appendix A: Soils Related Environmental Covariate Maps.

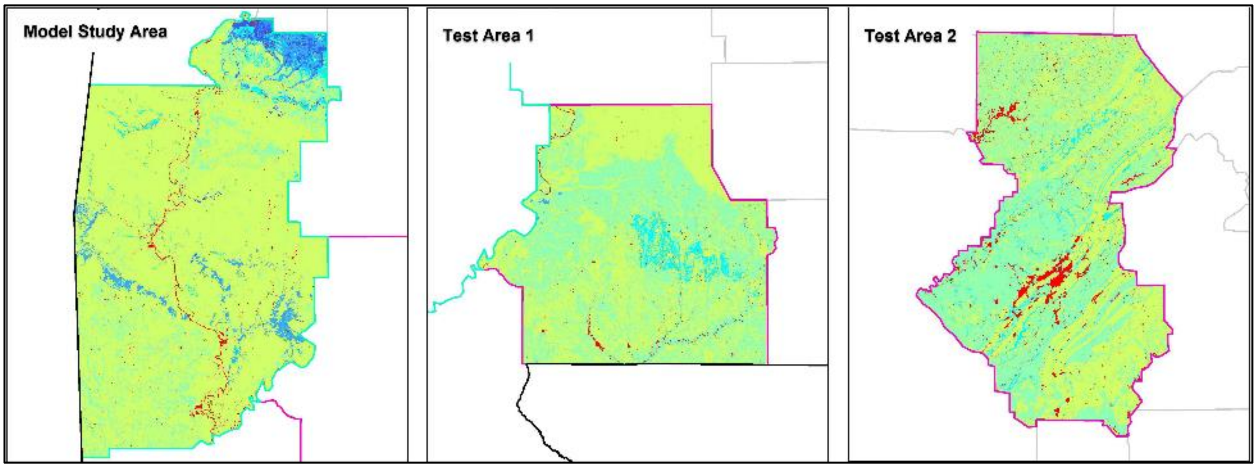


Figure 19: Soil pH thumbnail images for each of the study areas. See Appendix A for larger images.

Although the soils dataset from gSSURGO is provided in 30m grid cell size, the original soil mapping units were in vector format and were based on polygons with a minimum polygon map unit size ranging from one to ten acres (Ervin and Holly 2011; Soil Survey Staff 2011). Therefore, some reduction in granularity may occur when the Mapunit vector data was converted to raster format using the Polygon to Raster tool in the Conversion toolbox in ArcGIS 10.6 (Ervin and Holly 2011). This tool was used to produce the Soil Particle Size raster layer. All other soils related data layers were produced from the gSSURGO Soils Data Development Tools toolbox, gSSURGO Mapping Toolset, Create Soil Map tool for ArcGIS. Figure 20 shows the workflow used.

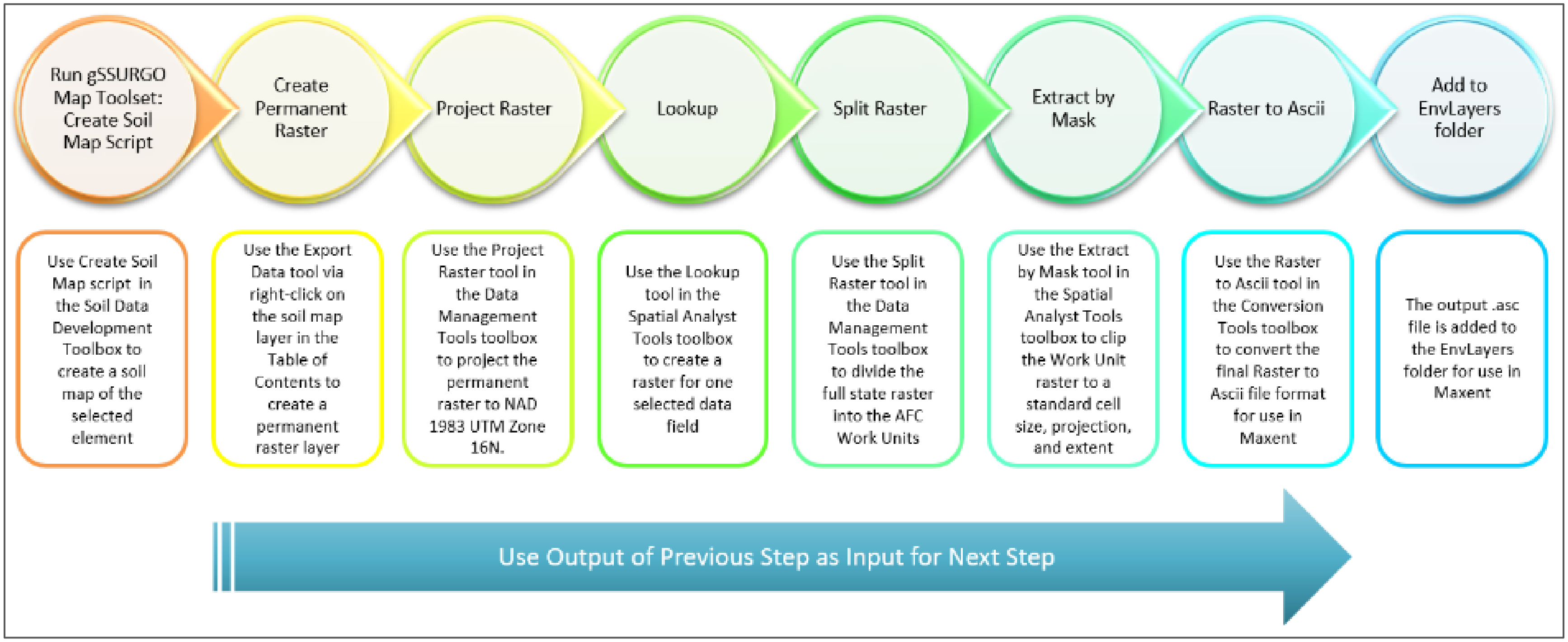


Figure 20: Data layer creation workflow for Soils data

3.3.3. Landcover Data Overview

Two data attributes from the Landcover dataset were included in the study. They include percent canopy cover and ecological system. Percent canopy cover for this study was pulled from the 2011 edition of the National Land Cover Dataset (NLCD) Tree Canopy cartographic layer produced by the Multi-Resolution Land Characteristics Consortium (MRLC). The 2011 edition was chosen as it most closely matched the timeframe that the initial Cogongrass infestation study was implemented and therefore would represent the percent canopy at the time of that study. The NLCD data is downloadable in 30m raster format and was generated by the United States Forest Service (USFS). Details related to the layer and its original creation can be found on the MRLC website (MRLC accessed 04/13/2019). This layer followed the preparation process as depicted in Figure 21.

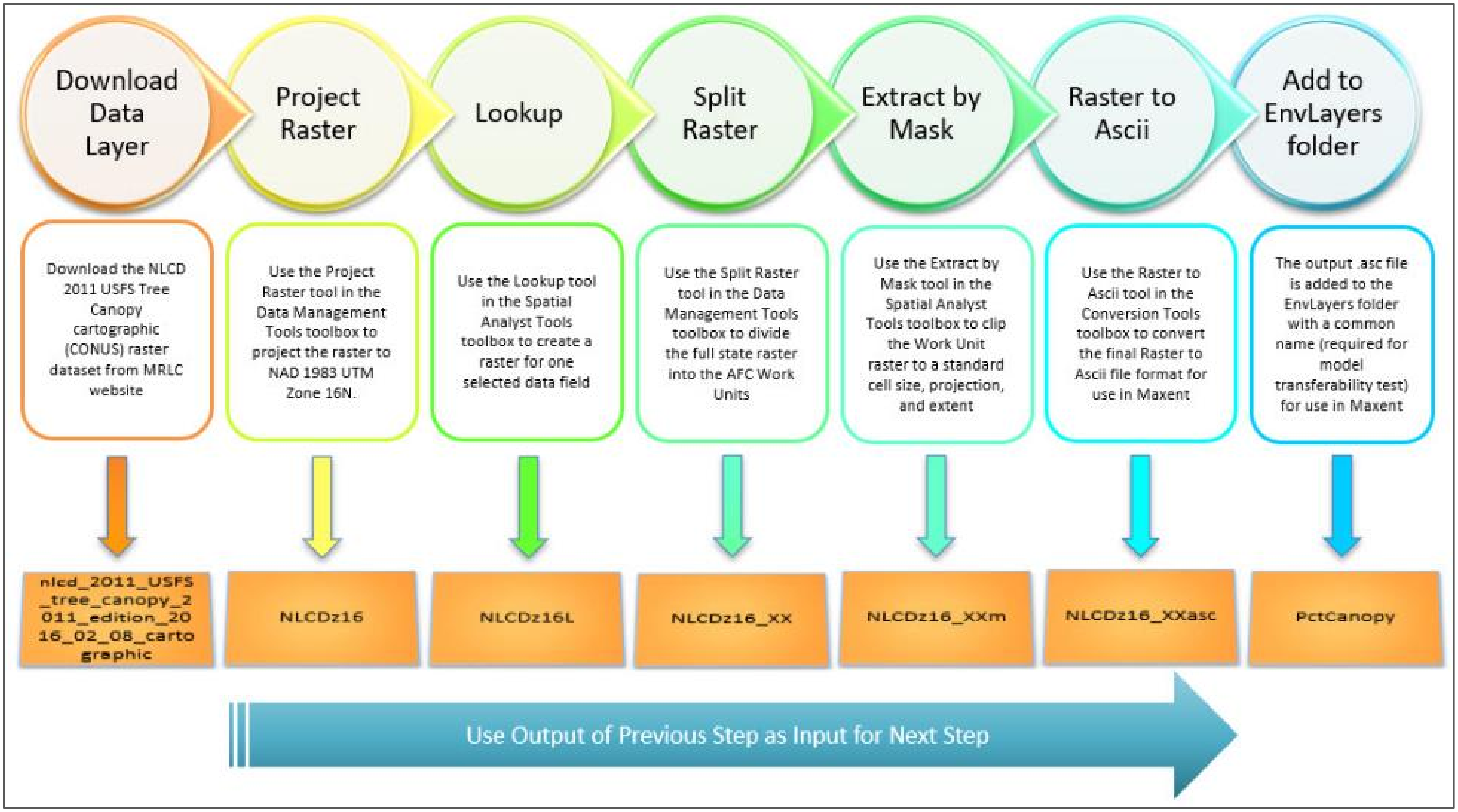


Figure 21: Data preparation workflow for Percent Canopy layer.

The percent canopy cover (PctCanopy) dataset was chosen for this study as several studies have suggested that canopy cover is a limiting factor in cogongrass growth as the species is somewhat shade intolerant. Percent canopy cover had a 77% contribution in the Mississippi portion of the Ervin and Holly study (Ervin and Holly 2011) and ability to survive as an understory species (Gaffney 1996) and tolerance up to 50% reduction in sunlight (Patterson 1980) has been reported. Figure 22 shows thumbnail images of this layer. PctCanopy ranges across the study areas from 0% to 100% where darker color indicates higher percentage. Large images of these layers are included in Appendix B: Other Environmental Covariate Maps.

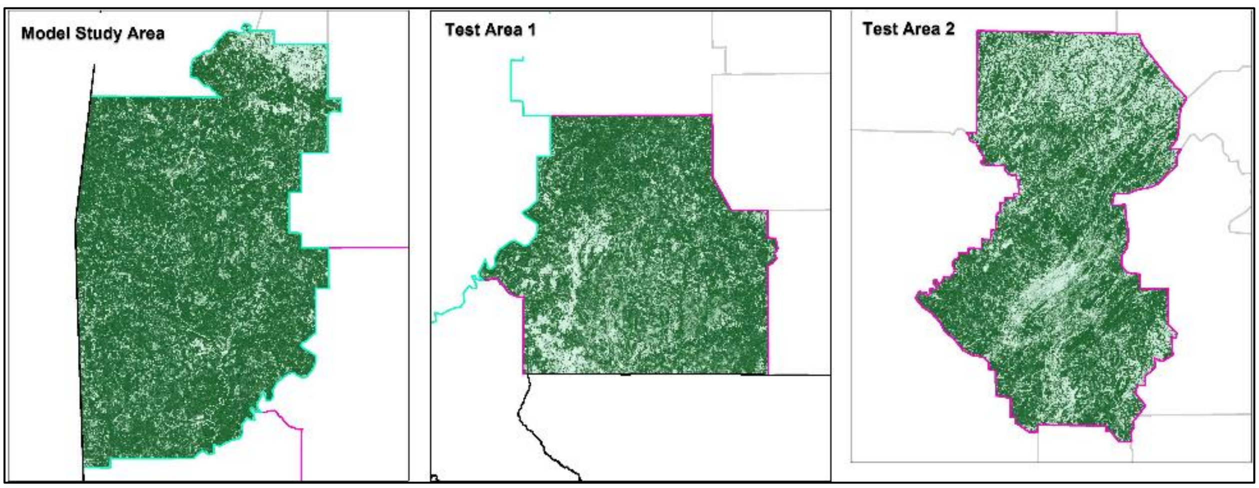


Figure 22: Percent canopy thumbnail images for each study area. See Appendix B for larger images.

Cogongrass can survive in a broad range of environmental ecological habitats as discussed in Chapter 1 and Chapter 2. Studies have also shown that cogongrass is particularly destructive in agricultural systems where the species can compete directly with agricultural crops thus creating not only an ecological impact but an economic one as well (Ervin and Holly 2011; Akobundu and Ekeleme 2000; Terry et al. 1997; Hubbard et al. 1944). To test the importance of ecological system on cogongrass infestation, the ecological system layer was generated from the GAP Land Cover Data for Alabama, USA (gap_30m_al). The method for creating the raster layer used in Maxent followed the same process as that used for the percent canopy layer described in Figure 21. Figure 23 shows thumbnail images of this layer. Ecological system is depcted in these images in grouped categories of Frest/Woodlands (green), Floodplain Forest (blue-green), Agriculture (tan), Developed (dark orange), Disturbed (brown), water (blue) and undefined/other (grey). Note the larger proportion of Floodplain Forest in the Model Study Area and the larger propotion of Developed land in Test Area 2. Large images of these layers are included in Appendix B: Other Environmental Covariate Maps.

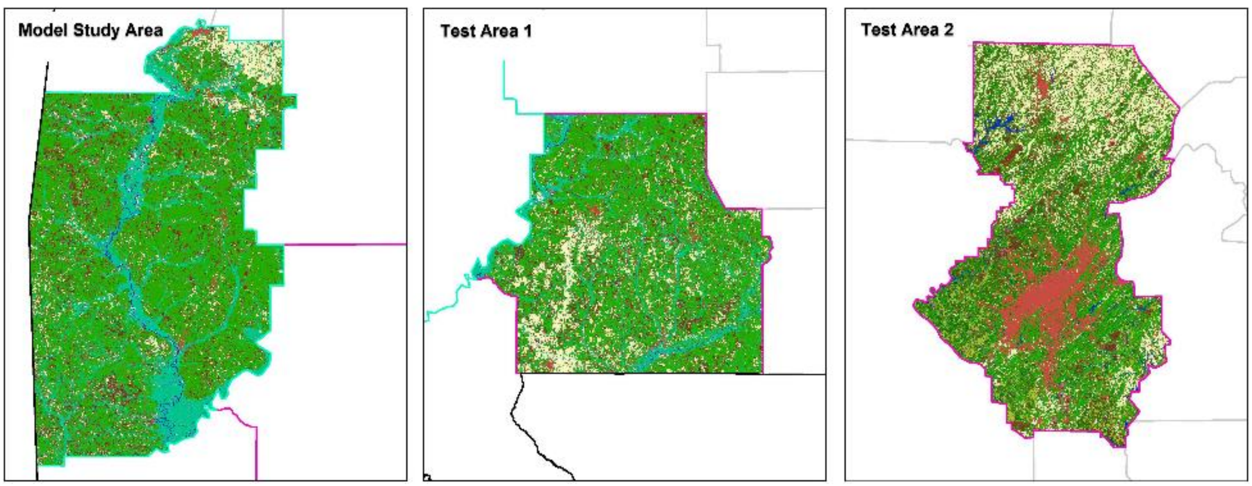


Figure 23: Ecological System maps for each study area. See Appendix B for larger images.

3.3.4. Roads Data Overview

Cogongrass spread occurs via two mechanisms, rhizome growth for local spread, and seed dispersal for long range spread as described in Chapter 2. Studies have indicated that spread along roads occurs due to wind dispersal as well as seed dispersal via hitch hiking on road maintenance equipment (Rauschert, Mortensen and Bloser 2017; Wilcut et al. 1988a; Wilcut et al. 1988b; Willard 1990). Although studies have quoted wind as the primary long-distance dispersal method (Yager, Miller, and Jones 2011), Willard suggests in his 1990 study that long range spread in Florida was primarily due to rhizome pieces being transported in fill dirt (Willard et al. 1990). In either case, roads play a part in infestation spread. To test for the impact of roads on Cogongrass, roads datasets were procured and the distance from each grid cell within a distance raster to the nearest road feature was calculated using the Euclidean distance tool in the Spatial Analyst toolset in ArcGIS 10.6. Roads data was provided for use in this study by Silvics Solutions, LLC in the form of four distinct road vector polyline layers. These layers were provided in North American Albers Equal Area Conic projection and were projected to NAD 1983 UTM Zone 16N using the Project tool in ArcGIS 10.6. The original roads layers provided consisted of a Local Roads layer, which contained both city and county roads, a State Highway layer, a US Highway layer, and an Interstate layer. Some feature overlap occurred between layers as some road features are captured in more than one dataset. This was mitigated when all road layer features were combined into a single layer and duplicates were removed.

The Euclidean Distance tool creates a raster dataset where each cell within the layer contains a value equal to the distance from the cell center to the nearest road feature. The use of the Near tool in the Proximity toolset was also investigated, but it was determined that the Euclidean Distance tool provided an output that best meet the needs of the Maxent model. The Near tool was utilized in data review, however. The input data layer for the Euclidean Distance tool was set to the consolidated roads layer. All four of the original roads datasets were combined into one consolidated road data layer in order to run the Euclidean distance tool on a raster layer depicting all roads at once. Figure 24 shows thumbnail images of this layer. The distance from nearest road ranges from 0 to 5359 meters across the three study areas. Note that in Test Area 2 there are substantially more local road features and therefore fewer cogongrass presence points that fall at great distance from roads. Large images of these layers are included in Appendix B: Other Environmental Covariate Maps

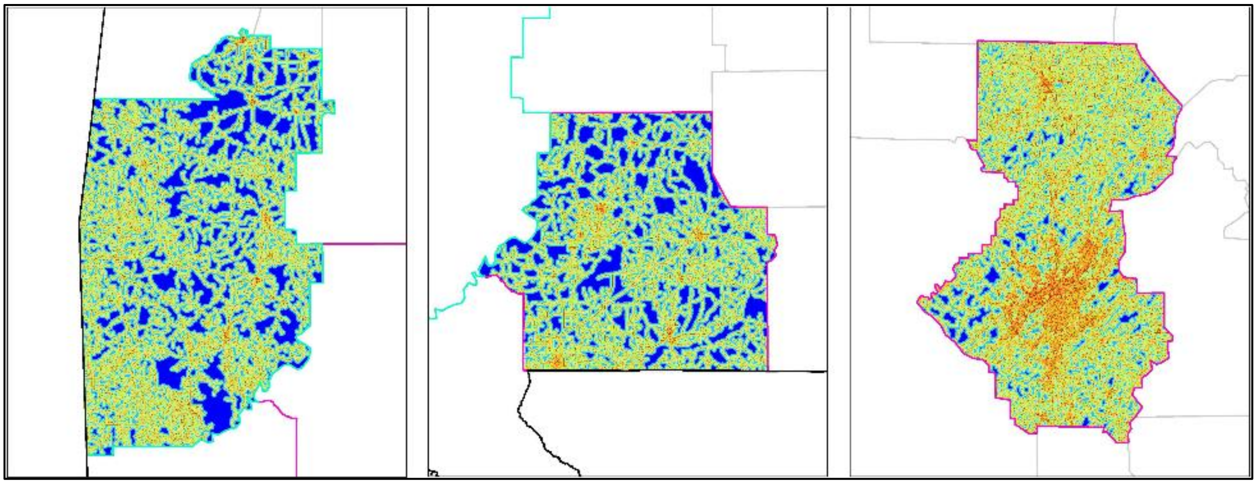


Figure 24: Distance to nearest road data maps for each study area. See Appendix B for larger images.

There are two important notes regarding the Euclidean distance raster. First, road width was not considered when this road layer was created as it was created from vector polyline layers that was then transformed into a 30m raster layer using the Polygon to Raster tool in the Conversion toolbox in ArcGIS 10.6. As previously mentioned, according to Ervin and Holly, the Polygon to Raster tool can result in some reduction in granularity (Ervin and Holly 2011). The output distance is to the center point of the road not the road edge. Second, when species presence points of larger distances from roads were visually investigated (using the Near tool and measure tool), many of these cells were within closer proximity to unmapped roads, such as interior woods roads, than is indicated in the Euclidean Distance raster layer which depends on mapped features being present in the dataset. Therefore, it may be a worthwhile endeavor to recreate this dataset in a later study with more granular roads data. This road dataset cleaning and augmentation is out of scope for this project.

3.4. Methods

Modeling methods for this study were loosely guided by the Ervin and Holly (2011) study in which the authors tested the transferability of a Maxent model developed for cogongrass location point data collected in the De Soto National Forest (NF) and Sandhill Crane National Wildlife Refuge (NWR) areas in southeastern Mississippi to a site consisting primarily of commercially managed pine timberlands in Clarke County, AL. This study piqued my interest in transferability of Maxent models and prompted a more Alabama centric study of transferability.

3.4.1. Defining the Model

Detailed review of biologic and ecological requirements of cogongrass was conducted to determine what environmental layers should be considered for this study. Ervin and Holly's (2011) study included soils related variables containing available water capacity, effective cation exchange capacity, percent organic matter, pH, and percent Silt content; and Land Cover related variables including percent canopy cover, and percent by ecological system (agriculture, coniferous forest, deciduous forest, developed, harvested forest, managed forest, and other). In the current study, available water capacity (AWC), effective cation exchange capacity (ECEC), and percent organic matter (PctOM) were not used.

AWC was discarded for two reasons, the first being insight gained from the 2000 study by King and Grace examined soil moisture content's effect on cogongrass seedling germination and growth and second, because in initial test run iterations AWC added little to no gain when reviewing the Jackknife results of preliminary default Maxent runs.

ECEC was not included as it was a surrogate for total soil nutrient content and availability in the Ervin and Holly (2011) study. For the current study, it was decided that this was too broad of a valuation metric and other studies reviewed have sited cogongrass' ability to tolerate a broad range of soil nutrient levels.

Percent Organic Matter was not used in the current study as the heuristic review of soil organic matter data within ArcGIS 10.6 revealed little difference in percent organic matter from 0 to 200cm of soil depth across the entire state of Alabama. This layer also added little to no gain when reviewing the Jackknife results of preliminary default Maxent runs. The other environmental covariates used in the Ervin and Holly (2011) study were included in the current study as well as the addition of percent clay, percent sand, particle size, and distance to nearest road.

The Maxent model was first run with the default settings in place as a baseline of model fitness for use and to assist in the determination of which parameters would need to be tuned in order to fit the model for the species and study location. Five replicates were run for the Model Study Area in Maxent for the tuned model followed by five replicates each run against the test areas with species and environmental layers masked to the Model Study Area using the Projection layers directory/file setting.

3.4.2. Tuning the Model

Although Maxent default settings were determined by Phillips based on testing across a wide range of species and environmental factors, it is suggested that models be tuned for the specific species and location being modeled (Elith et al. 2011; Phillips 2017). Model tuning was performed to maximize performance while minimizing the potential for overfitting. The Model Study Area Maxent model was built using the biological environmental variables relevant to AFC Work Unit 11 and Cogongrass in general (Table 6). Care was taken to select environmental variables that both represented specific measures relevant to the biology and habitat preferences of the species and were broad enough to be useful measures across the landscape. This same thought process was given to tuning the model.

Several Maxent default settings were maintained in this study. The regularization multiplier was left at 1 as was the case in the Ervin and Holly study. This value is a modifier to help smooth the model in an attempt to avoid over-fitting and underfitting and helps to balance fit and complexity within the model (Ethel et al 2010). Modifying this value was tested with settings of 0.5, 0.8, and 2 (Table 8) with limited improvement when the regularization multiplier was reduced and limited reduction in fitness when the regularization multiplier was doubled.

Table 8: Regularization Multiplier's effect on AUC. All other settings remaining equal.

Regularization Multiplier	AUC (Training/Test)
0.5	0.712/0.717
0.8	0.709/0.715
1	0.708/0.715
2	0.699/0.705

The AUC calculation on one replicate run was used as the indicator of fitness while running tuning tests. The decision to leave the default setting for these values was made as the change in AUC due to modification of regularization multiplier alone did not significantly change the modeled results. The number of background points, maximum number of iterations per replicate run, convergence threshold, and default prevalence were all also left at their default settings.

Parameters that were tuned in the model included values that resulted in modification of the model itself and values that resulted in modification of the output from the model. Parameters that resulted in modification of the model itself and were tuned in the Model Study Area Maxent model were changing the output format to Logistic, modifying the replicate run type, setting the number of replicates, selecting to add samples to the background, and selecting to use samples with some missing data. Parameters that resulted in modification to the output files of the model but not the model itself included: selecting to create response curves and run jackknife tests, increasing the number of processor threads used by the model, selecting to write plot data, selecting to add summary results to the Maxentresults.csv file, and selecting to write background predictions. Appendix D: Maxent Model Settings Screen Captures shows how all of these were set within Maxent. Some of these modifications are discussed below.

The Logistic output was selected rather than the newer default of Coglog as Logistic output was the default in previous versions of Maxent and was the output format selected by Ervin and Holly. Logistic output is also recommended in Phillips and Dudik 2007. Modifying this setting increased the AUC of the resultant model marginally (0.698 to 0.708) but this minor difference is most likely due to the nature of the random seed setting. Increasing the number of processor threads allowed the model to use more of the computer's processing capabilities thus allowing some intensive processes such as jackknife creation to run faster. Checking the setting to write background predictions was required in order to calculate TSS for each model replicate run. Modifying the replicate run type involved setting the replicate run type to sub sample along with setting the random test percentage to 50% and checking the random seed checkbox. These three settings in conjunction provided a slightly better model AUC than using the default Cross validate replicate run type (AUC improved from 0.708 to 0.725). The final model selected to use in the study was the model with AUC of 0.725.

3.4.3. Gauging Fitness of the Model

The purpose of evaluating a model is to determine if it is useful, or fit, for the purpose the model is being used for (O'Sullivan and Perry 2013). The Maxent model uses parameters and constraints to modify the model output. Many studies using Maxent modify a few key parameters but leave most parameters set to their default values. Depending on the study subject, this may be an appropriate course of action. An analysis of the parameters and settings needed to produce an appropriate model was performed in this analysis to ensure that model parameters were appropriate to the study.

The fitness of the model in predicting the distribution of cogongrass was evaluated using AUC, Sensitivity (Omission) and TSS. The relative contribution of each environmental variable to the model was evaluated by review of the jackknife output tables as well as the plot graphs of each individual environmental variable. In Section 3.4.1, analysis of these metrics allowed for the removal of datasets that proved to be of little value to the study.

The Maxent model was set to five sub-sample replicates, withholding a randomly selected 50% of the test data in each iteration. Setting the model to 10 replicates was also tested however the statistics were not significantly different between the five replicate and 10 replicate tests. The model results reported in Chapter 4 represent the resultant predicted potential distribution of cogongrass given suitable conditions, averaged across five model replicates as well as an averaged standard deviation.

3.4.4. Testing Transferability of the Model

To evaluate the effectiveness of the trained model, the predicted distributions produced with new environmental data for the test areas were compared to additional verified presence points of cogongrass locations in the test areas (test data). These test data were used to determine the accuracy of the model's predicted distribution, and therefore the viability of the model in predicting distribution for cogongrass infestations when transferred to different geographic space than where the model was originally trained. A model that was trained on a set of environmental variables in one geographic space can be transferred by running the same model using the same set of environmental variables that have been extracted to the new study area location (Phillips et al. 2017).

When the model is transferred to a new geographic area, the Projection layers directory/field in the Maxent user interface can be set to point the new model run (in this case model runs for Test Area 1 and Test Area 2) to the original environmental layers (in this case the Model Study Area's environmental layers) so that the environmental layers used in the new model runs are "clamped" to the original model layer ranges. Clamping essentially sets any layer value in the new model run's environmental layers that fall outside of the range of values for that layer in the original model run to equal the outer bound of the original environmental layer. For example, the range of percent silt in the PctSilt environmental layer for the Model Study Area was 0 to 60%, for Test Area 1 the range was 0 to 56.1% and for Test Area 2 the range was 0 to 66%. Therefore, the new model run for Test Area 1 did not require clamping to the extent of the range from the Model Study Area for the environmental variable, but PctSilt did require clamping for the new model run for Test Area 2. The response to this variable in the new model run for Test Area 2 is held constant for all values that fall outside of the training range (the range of values found in this layer for the Model Study Area model run) essentially treating those values as if they were at the limit of the range (in this case, 60%). According to Phillips in the updated (2017) tutorial on Maxent, testing transferability by projecting the model in this manner is appropriate when the goal is to evaluate a model at a set of test locations (Phillips 2017) which is the goal of this study.

In this analysis, transferability of the model was tested by utilizing regional subsets of the same environmental layers and maintaining the same parameters in both of the test areas. During the initial environmental data layer creation process, the AFC Work Unit polygons were used as boundary extent to split the original data layers into 18 separate raster layers. By using the Split Raster tool in ArcGIS 10.6, transferability testing was substantially sped up, as data layer manipulation requirements were lessened.

Maxent was run against Test Area 1 and Test Area 2 utilizing the same setting parameters as were defined in the Model Study Area's Maxent model. For each of the test areas, the presence points .csv file was created by extracting only those presence point feature that fell within the boundary of the selected AFC Work Unit. This file was then processed and converted for use in Maxent as defined in Section 3.3. All environmental layers utilized in the study were also extracted to the extent of both test areas as separate files following the same processes as defined Section 3.3. The environmental layers folder was set to the folder housing the .asc files for the test area included in the transferability test model and the output directory was set to the test area folder's output directory. The projection layers directory/file was set to the directory that housed the environmental layers used in the original Model Study Area model run.

It is important to note that all layers, in each of the environmental layers directories should use a common naming convention so that the Maxent model can determine appropriate layer clamping for the Test Area model runs. For example, the drainage class layer is named "DC" in all three directories (Modeled Study Area, Test 1, and Test 2). The model is then trained using the environmental layers set in the environmental layers list for the Model Study Area model and then transferred onto the Test Area environmental layers which clamps the environmental variables for the current run to the bounds of the original model run to which the layers are being transferred (Phillips 2017).

TSS is a special case of Kappa that reduces issues associated with prevalence that prevents Kappa from being a useful metric for presence only data. TSS was used as a measure of model fitness in this study. The formula for calculating TSS is shown in Equation 1.

$$\text{TSS} = \text{Sensitivity} + \text{Specificity} - 1 \tag{1}$$

A single TSS score for each model was determined by calculating the TSS for each of the five replicates in a model run individually and selecting the run with the highest TSS score to be the representative score for that test. TSS can be calculated from the Maxent output by selecting the “write background predictions” selection on the Experimental tabs in Settings. This setting tells Maxent to write background predictions files for each of the replicate runs.

Next, copy the “logistic” column from the background predictions file of replicate 0 and paste it into column A of a spreadsheet (tab labeled 0). Then, open the sample predictions file for replicate 0 and copy the “logistic prediction” column into column B of your spreadsheet. Step three is to choose which threshold you want to use in the calculation. In this study I have chosen to use the 10 percentile training presence logistic threshold. This value can be found in the “Maxent Results” file as output by Maxent. For step 4, do a count of sample predictions test values greater than the threshold for replicate 0, a count of sample predictions test values less than the threshold of replicate 0, a count of background prediction test values greater than the threshold for replicate 0, and a count of background prediction test values less than the threshold for replicate 0. With these values in hand TSS for replicate 0 can be calculated.

Calculate Sensitivity as the count of cells where the sample predictions test values are greater than the threshold, divided by the total count of sample prediction test values. Then, calculate Specificity as the count of cells where the background predictions value is greater than the threshold, divided by the total count of background prediction values. As defined above, the

TSS for replicate one is Sensitivity plus specificity minus one. This calculation is repeated for each replicate in the Maxent run and the largest TSS from the replicate set was then used as the TSS score for the model in this analysis.

Chapter 4 Results

Studies exploring the application of Maxent have indicated that there is no perfect metric to evaluate all models for fitness to the study. Species such as cogongrass, which can tolerate a broad range of ecological and environmental conditions, can produce model results with a large area of predicted occurrence (Ervin and Holly 2011). It is suggested that each evaluation metric be assessed in context with the specific species and variables in use and the desired use of the model output. It is also suggested that a mix of evaluation metrics be used to determine model suitability and fitness (Anderson 2012; Merrow, Smith, and Silander 2013; Radosavljevic and Anderson 2014; Peterson et al. 2011). To this end, AUC, Sensitivity (Omission Rate), and TSS were selected as measures of model suitability.

The Model Study Area Maxent model used in this analysis was evaluated using a 5-fold sub-sample with 50% of the presence points set aside randomly for testing the model. It was appropriate to use 50% of the presence points for testing due to the large number of presence points included in the species presence data layer. The Model Study Area Maxent model results were compared to the results from the Test Area 1 Maxent model and Test Area 2 Maxent model transferability tests and the model suitability results for these three models (Model Study Area Maxent model, Test Area 1 Maxent model, and Test Area 2 Maxent model) utilized in the study are shown in Table 9and are discussed in greater detail below.

Table 9: Model Suitability Indicator Results

Indicator	Model Study Area	Test Area 1	Test Area 2
AUC (5 fold sub-sample)	0.7250	0.7460	0.8460
AUC std dev	0.0010	0.0020	0.0170
Test Omission	0.0832	0.0807	0.2941
TSS (highest of replicates)	0.4087	0.3944	0.2377

4.1. Area Under the Receiver Operating Characteristic Curve (AUC)

The area under the receiver operating characteristic curve (AUC) indicates fitness of the model (Phillips et al. 2017). As discussed in Chapter 2, the AUC shows the average sensitivity vs. specificity for the species being modeled and tells us how well the model can discriminate between presence locations and background data. According to Elith et al. (2011), an AUC of 0.70 and above indicates sufficient fit for ecological niche study purposes. Therefore, the AUC of 0.725 returned for the Model Study Area, along with its low standard deviation (0.0010), indicates a stable model and a good fit for predicting the distribution of cogongrass within the study area. Both of the transferability test areas returned AUC above 0.70, however the much larger standard deviation in Test Area 2 warrants some concern.

The mean standard deviation for the AUC in the Model Study Area is 0.0010, which is very low and a good indication of model stability. The AUC of Test Area 1 (0.746) is slightly higher than that of the Model Study Area (0.725) with similarly low standard deviation (0.0020). This is an indication of good transferability of the model to Test Area 1. The AUC of Test Area 2 (0.846) is even higher than that of Test Area 1, however the standard deviation is higher at 0.017. Although this standard deviation is still within a valid range (95% of the replicate runs fall within one standard deviation) it is much higher than the standard deviation of the AUC for the Model Study Area and that of Test Area 1. Therefore, additional review of the results of the model in Test Area 2 is required.

4.2. Sensitivity (Omission)

The omission rate for the model is depicted in the .html output file produced by Maxent. The omission rates for test samples from the model run on each of the three study areas are provided in Table 9. The omission rate shows model performance as a function of the predicted occurrence. For the Model Study Area and Test Area 1, the modeled omission rate follows the predicted omission closely with very low standard deviation. The Model Study Area omission rate falls at 0.0832 and Test Area 1's omission rate similarly falls at 0.0807. This shows a very good match of the test data to the trained model predictions and is an indicator of a well fitted model. For Test Area 2 the Omission rate was significantly higher (0.2941) again, warranting a closer look.

4.3. Variable Contributions and Gain

An understanding of how the environmental variables selected for use within the model effect the model outcome is important in understanding the statistics used to test model fitness. Maxent produces very detailed output in the form of multipage html documents. In these documents, tables showing the percent contribution and permutation importance of each individual environmental variable included in the study assists in understanding the model results. Also, in these documents, response curves provide a visual representation of the predicted potential distribution of species occurrence in two graphs per variable.

The percent contribution indicates how much the individual variable contributes to the fit of the model (gain). This value should be used with caution when variables are highly correlated (Phillips et al. 2007; Phillips 2017) which is a potential issue with the particle size dataset used in this analysis. The permutation importance shows the contribution of each variable via random permutation (and does not rely on the path the model used to get to the final result) thus a larger permutation importance indicates that the model depends heavily on the variable. For this analysis we focus on the permutation importance of variables, as this value lessens the impact of variable correlation. The percent contribution (Table 10) and permutation importance (Table 11) for the Modeled Study Area as well as Test Area 1 and Test Area 2 are provided below.

Table 10: Percent contribution of environmental variables for the Model Study Area, Test Area 1, and Test Area 2.

Environmental Variable	Model Study Area	Test Area 1	Test Area 2
PctCanopy	61.1	38.7	20.1
EcolSys	11	7.7	9.4
Distance	10	2.1	56.4
pH	8.2	0.3	0.2
PartSize	3.5	19.3	1.1
DC	3.3	17.7	1.2
Bed	1.1	0.6	7.1
PctClay	0.7	3.2	1.3
PctSand	0.5	5.5	2.9
PctSilt	0.6	5	0.3

Table 11: Permutation importance of environmental variables for the Model Study Area, Test Area 1, and Test Area 2.

Environmental Variable	Model Study Area	Test Area 1	Test Area 2
PctCanopy	56.9	30	36.6
EcolSys	14.7	8.8	4.1
Distance	8	2.4	43.5
pH	8.2	1.1	0.1
PartSize	4.4	16.3	2.3
DC	3	1.3	2.5
Bed	1.1	0.9	7.5
PctClay	0.7	13.9	1.5
PctSand	2.2	8.6	1.2
PctSilt	0.8	16.8	0.7

Review of individual environmental variable's response curves in conjunction with data on percent contribution and permutation importance provides valuable information into the impacts that each variable has on the model outcome. The response curves for each environmental variable included in the model (Appendix E: Response Curves) indicate how each variable affects the predicted probability of presence when all other variables are set to their average value. This means that each curve shows the marginal impact of change to the predicted potential distribution of cogongrass infestation given suitable conditions resulting from changing just the one variable selected. Each curve shows the mean response of the 5-fold sub-set model in red and +/- one standard deviation in blue. Below the four most significant variable's response curves for each model are explored.

In keeping with previous studies (Ervin and Holly 2011), percent canopy cover (PctCanopy) had the highest permutation importance for both the Model Study Area (56.9%) and Test Area 1 (30%) and fell to second highest for Test Area 2 (36.6%) (Figure 25). This is as expected as cogongrass is somewhat intolerant to shade, and previous studies have shown shade to be an important factor when predicting the distribution of the species (Ervin and Holly 2011; Gaffney 1996; Patterson 1980). This is also consistent with the Ervin and Holly (2011) study in which PctCanopy had a relative contribution of 77% (recall that their study focused on heavily forested ecosystems in Mississippi). As the current study hypothesized, it was expected that PctCanopy cover would have significant impact on predicted site suitability.

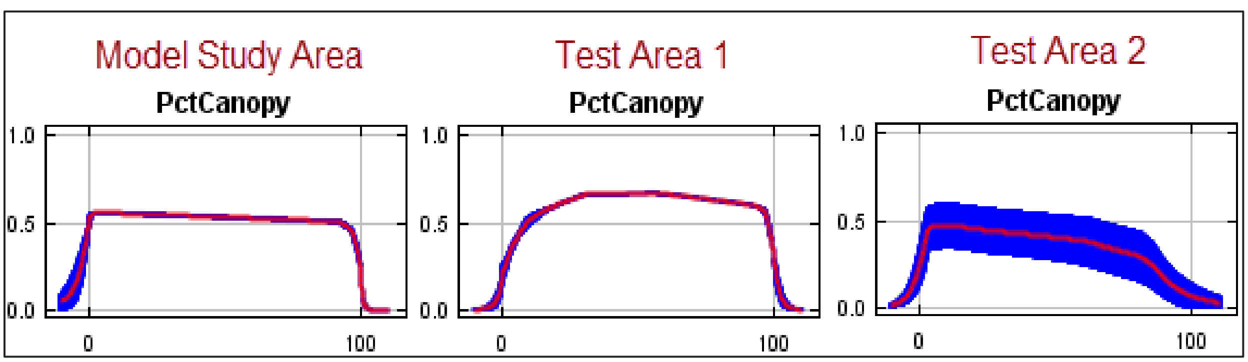


Figure 25: Response Curves for percent canopy.

It is relevant to observe that in Test Area 2, where there is significantly more developed and open (agricultural) land, the permutation importance of PctCanopy was lower than in the other two model areas. As can be seen in the PctCanopy graphs above for both the Model Study Area and Test Area 1, the impact of PctCanopy on the model remains relatively high and consistent with low standard deviation (thin blue area around the red response curve). The impact of PctCanopy for Test Area 2 was much more variable across the replicate runs as indicated by the thick blue area around the red response curve for the Test Area 2 graph. The response curves for PctCanopy for all three models indicate a high impact of this variable on each model as the curves all increase exponentially at the beginning of the range and decrease just as dramatically at the end of the range.

Ecological System (Eco) was the second highest permutation importance in the Modeled Study Area (14.7%). For Test Area 1, Eco was not amongst the top four in permutation importance (8.8%) and for Test Area 2, eco was the fourth highest in permutation importance (4.1%). This potentially shows some departure in consistency and thus transferability where ecological system is concerned. Ecological system is a categorical data set and the importance of each individual category plus or minus one standard deviation to the averaged marginal response of the model to changing one variable is shown in Figure 26. In the graphs in Figure 26, the missing columns represent ecological systems that have no impact on the potential distribution in the model indicated. See Appendix F for definition of the categorical values for ecological system for each model. Most ecological systems have average impact on the model (hovering around 0.5 for the Model Study Area and Test Area 1 where over half of the systems have less impact in Test Area 2 and the remainder's impact is more volatile.

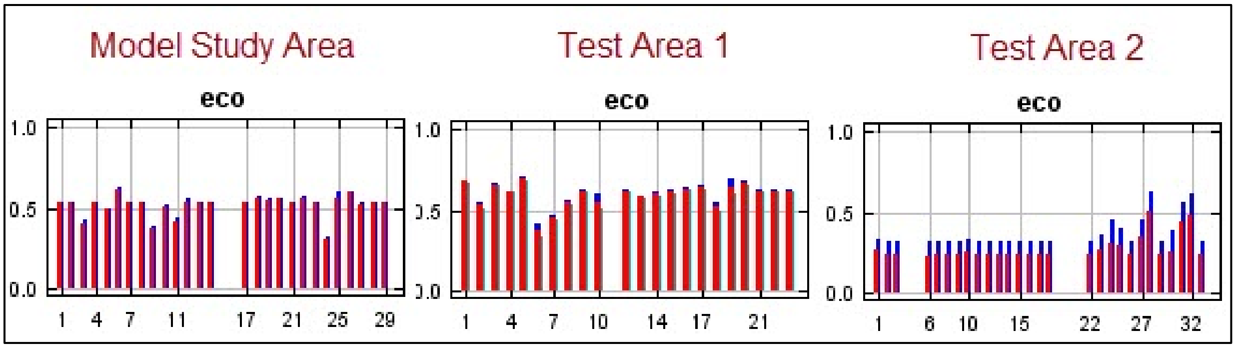


Figure 26: Response Curves for Ecological System

At first review of the Maxent output response curves for eco, it is unclear which ecological systems have impact and which do not. In this instance it would be prudent to group the data results from the three models into one graph to better review the response of cogongrass to ecological system across the Model Study Area and the two transferability test areas. Figure 27 provides a graph of cogongrass' response to ecological system in each model with consistent numbering for each ecological system present.

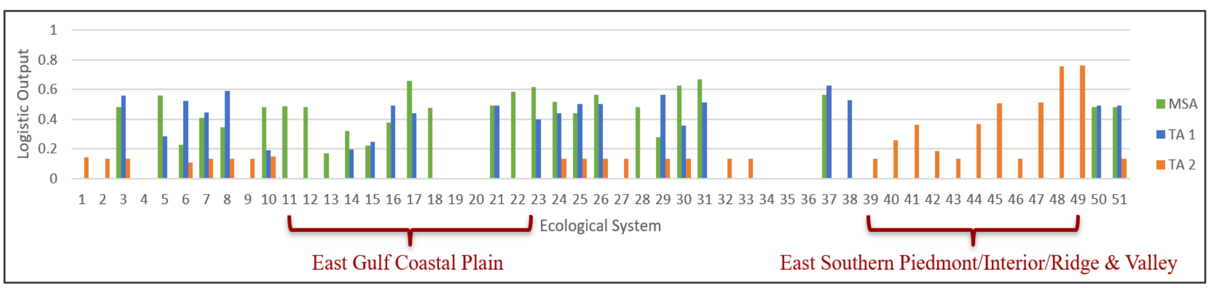


Figure 27: Consolidated graph of cogongrass' response to ecological system for each of the three models.

This consolidated graph shows that not all ecological systems are present in all study areas included in this test. Two such groups of ecological system are identified in Figure 27. Note that ecological systems 11 through 23 (East Gulf Coastal Plain ecological systems) do not have any orange bars associated with them indicating that these ecological systems are not present in Test Area 2. Also note that ecological systems 39 through 49 (East Southern Piedmont, East Southern Interior, and East Southern Ridge and Valley ecological systems) do not have any blue or green bars associated with them indicating that these ecological systems are not present in the Model Study Area or Test Area 1. These results suggest that the attribute granularity of this layer may be too detailed for this study. Table 12 offers a potential grouping of the ecological system into broader categories that are easier to consume. It is recommended in any future studies utilizing this variable, that this data be grouped as indicated in Table 12 and the models re-run to better gauge impacts of this variable on the models.

Table 12: Percent geographic area (km^2) occupied by each grouped ecological system within the Model Study Area, Test Area 1, and Test Area 2.

Ecological System Group	Model Study Area % of area	Test Area 1 % of area	Test Area 2 % of area
Forest/Woodlands	49.99	46.88	52.34
Floodplain Forest	26.54	22.24	0.14
Agriculture	8.07	14.64	20.84
Developed	3.18	3.09	15.05
Disturbed	10.99	12.39	8.19
Water	1.18	0.68	2.80
Other	0.06	0.07	0.64

For the Modeled Study Area, pH was the third highest in permutation importance (8.2%) but did not rank in the top four for either of the two test areas. Cogongrass is tolerant of soils with a range of pH values but has been shown to grow best in relatively acidic soils (pH of 4.7) (Wilcut et al. 1888a). The graphs in Figure 28 show the impact that the pH covariate has on the predicted probability of presence given that all other variables are kept at their average value. The Model Study Area graph for pH shows a gradual increase in impact as pH increases from 0 to 8.3 and the impact decreases at pH values higher than 8.3. Test Area 1 exhibits a different pattern in pH's impact on the model. In Figure 28 pH graph for Test Area 1 shows that impact remains high but relatively static until it reaches 5.0 then the impact due to pH decreases slightly and the related standard deviation of impact increases as the pH approaches 8.0. For Test Area 2, the pH runs a similar curve to the Model Study Area but with much greater variability to impact between model run replicates.

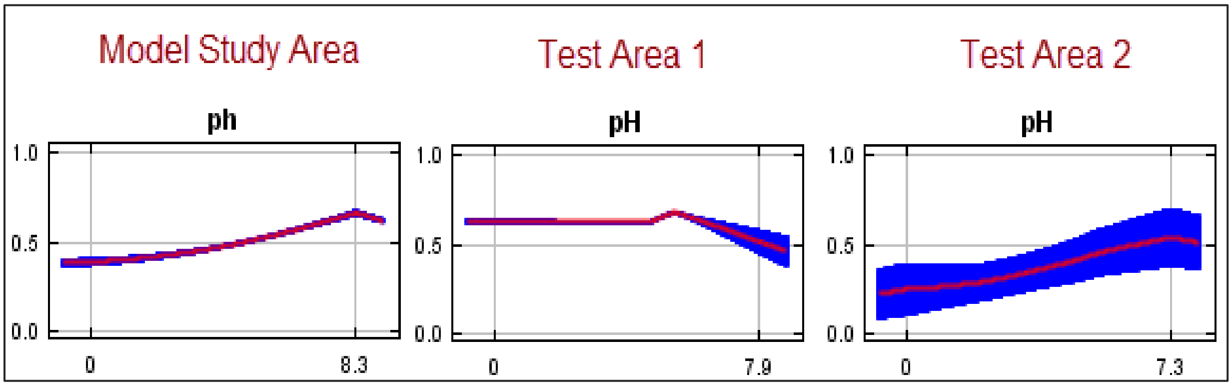


Figure 28: Response curves for pH

Distance to road was the fourth highest in permutation importance (8.0%) for the Model Study Area but did not rank in the top four in Test Area 1 and was the highest in permutation importance for Test Area 2. As would be expected given the spikelet wind dispersed travel distances discussed in the Yager, Miller, and Jones (2011) study briefly described in Section 2.3 and the physical seed dispersal distances as shown in the Rauschert, Mortensen and Bloser (2017) study, Figure 29 shows that distance to road has its greatest impact in close proximity to roads with stable to lessening impact as distance increases for the Model Study Area.

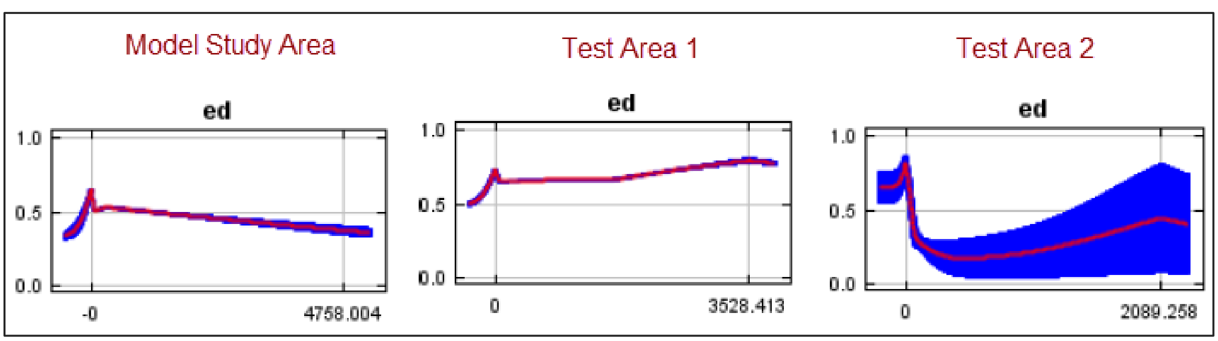


Figure 29: Response curves for distance to nearest road.

For Test Areas 1, impact to the model due to the presence of the distance to roads variable with all other variables remaining at their average rate followed a similar curve through roughly 2000 meters then increased in impact as distance increased. This could potentially be due to a lack of interior woods roads in the roads layer dataset as noted via visual inspection of the data in ArcGIS 10.6. For Test Area 2, the impact of distance to road was high at close proximity then dropped exponentially and recovered only minimally on average over the course of the range of distances for the layer. Test Area 2 also shows significant standard deviation (greater than 1) as the curve approaches its maximum distance values. As discussed in Section 3.3.4, the road dataset could be improved with added local roads information in non-urban areas. Test Area 2 contains significantly more local roads due to its higher percentage of developed land compared to the Model Study Area and Test Area 1. It is recommended that a spatial data creation project be launched to augment the roads dataset should this variable be considered for future study.

For Test Area 1, percent silt (16.8%), particle size (16.3%) and percent clay (13.9%) were the second, third, and fourth highest in permutation importance respectively in that model. None of these variables ranked in the top four in the Model Study Area. These three variables are correlated as the particle size categorical values are defined based on the soil texture percentages defined by the percent of silt, clay, and sand. It was determined that including both the categorical particle size variable and the component specific percentages was important to the analysis. Percent silt provided a percent contribution of 10% to the Ervin and Holly study.

For Test Area 2, depth to soil restrictive layer ranked third in permutation importance (7.2%) although this variable did not rank in the top four for either the Modeled Study Area or Test Area 1. As Test Area 2 was selected to test the transferability of the model because of its many dissimilarities to the Model Study Area, it was expected that this location would differ in variables of importance. For a more detailed discussion of how the test sites were selected, refer back to Section 3.1.

Finally, the jackknife tests run by Maxent are depicted in three charts in the output .html. For this analysis, we focus on the Jackknife of test gain charts to judge variable impact on the model as test data is used to judge model performance. The Jackknife of regularized training gain and Jackknife of AUC could also be used to test model performance. Since we are focused on testing the model and the model's transferability to new geographic regions within the state, it was determined that the best test would be to utilize the Jackknife of test gain. "Gain is closely related to deviance, a measure of goodness of fit used in generalized additive and generalized linear models" (Phillips 2017, 4). Data in these jackknife charts are normalized, so all study area jackknife charts can be compared.

Figure 30 shows the Model Study Area's jackknife chart. Two points of interest are highlighted here. As was shown in the variable contributions table (Table 10), PctCanopy had the highest percent contribution and highest permutation importance to the model in the Model Study Area. The jackknife analysis seconds this conclusion. In the Jackknife of test gain, the PctCanopy row shows that PctCanopy had the most significant information by itself about the suitability of the environment for the species (blue bar) and had the most significant total impact in the form of reduction in gain, when omitted from the analysis (red bar). Therefore, PctCanopy provides the most independently important information that cannot be explained by use of other variables included in the model.

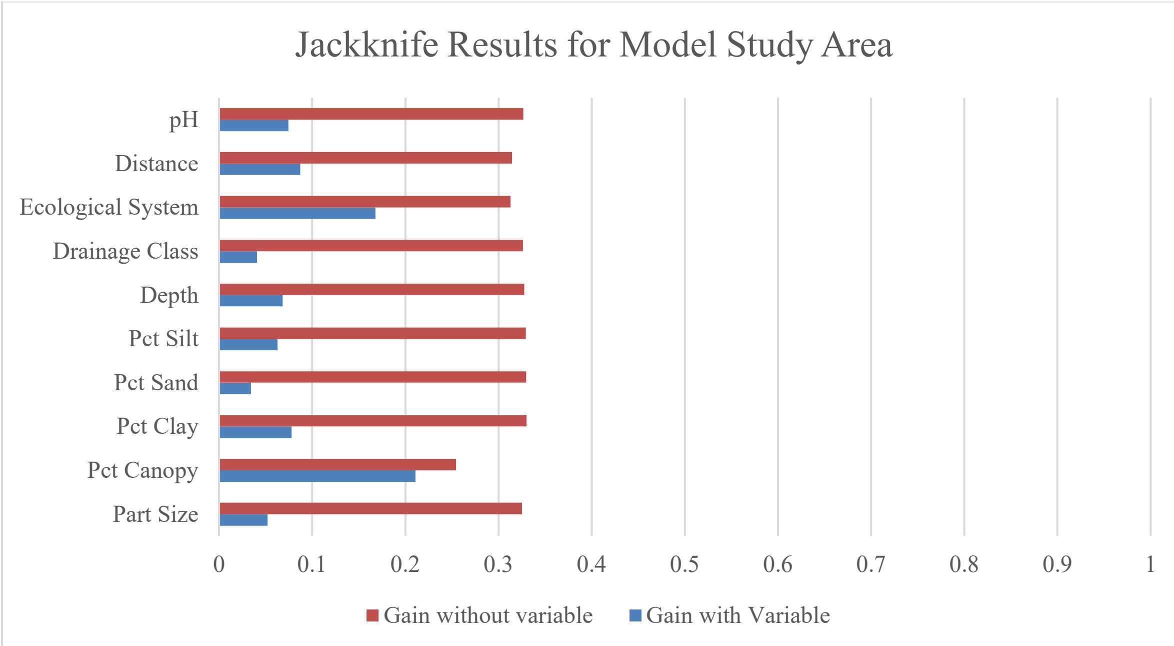


Figure 30: Jackknife of impact to gain by variables included within the model run for the Model Study Area.

For the test areas, the jackknife tests provide key information about the differences in variable impact on gain between the Model Study Area and the transferability test sites. Test Area 1, which was the most similar to the Model Study Area, also indicates, through the jackknife test, that PctCanopy has the most independently important information that cannot be explained by use of other variables included in the model just as was the case with the Model Study Area. It is important to note that the total percent contribution of PctCanopy droped for Test Area 1 which is in large part due to the increase in importance of other variables in the model outcome as seen in the jackknife results (Figure 31).

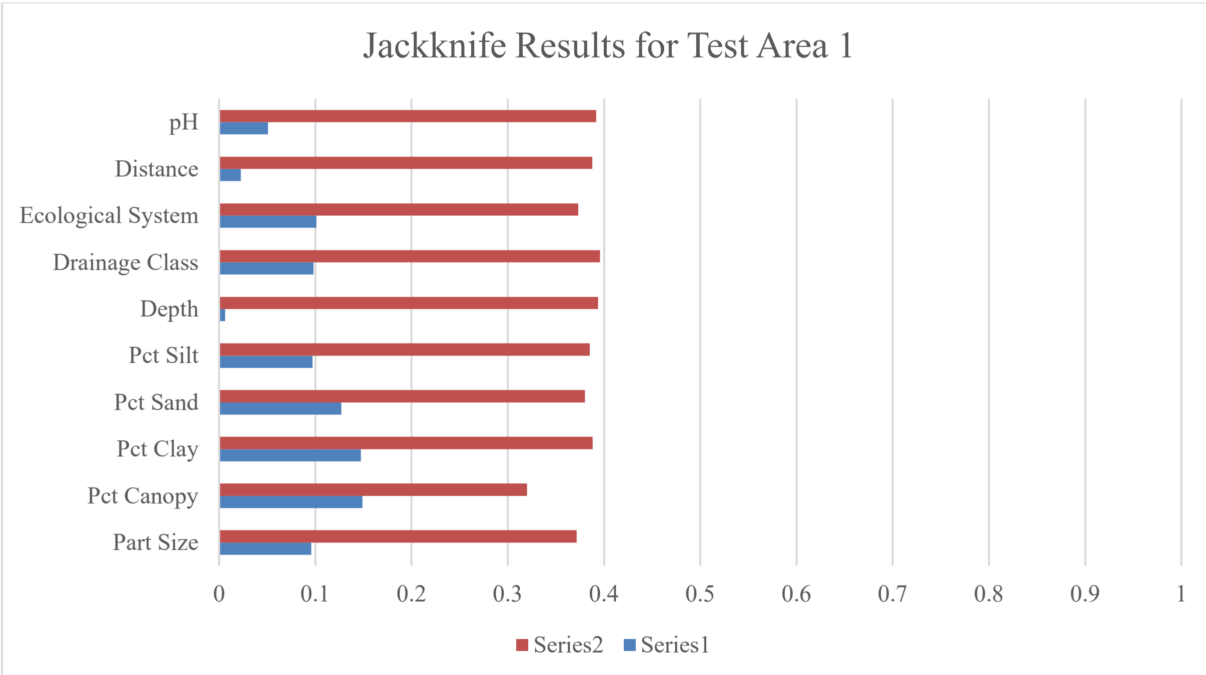


Figure 31: Jackknife of impact to gain by variables included within the model run for the transferability Test Area 1.

The percentages of the individual soil textures (clay, silt, and sand) each independently provide important information to the maxent model over the Modeled Study Area. The importance of ecological system dropped significantly between the Model Study Area model and the two test area models. Given the rural nature of Test Area 1 and the comparatively rural nature of the Model Study Area, it was expected that these two study areas would have similar variables of importance but from the results of the jackknife analysis, it is clear that soil texture plays a larger role in site suitability in Test Area 1 than it did in the Modeled Study Area with PctSilt providing 16.8% of permutation importance, PctClay providing 13.9% and PctSand providing 8.6% in Test Area 1. That being said, the inclusion of these layers in the original model was advantageous as this allowed for the transferred model in Test Area 1 to utilize these important factors. As discussed in Chapter 2, it is important for environmental variables selected for use in Maxent to be broad enough in biological or ecological extent yet specific enough to be valuable across the entire intended geographic region, when transferability of the model is a desired outcome. The AUC for Test Area 1 (0.746) shows that the model is fit for use in testing the probability of occurrence of cogongrass in this area.

The jackknife chart for Test Area 2 is quite interesting (Figure 32). It was expected that the model environmental layers for the Model Study Area would differ significantly in results from Test Area 2 as these two areas are quite different in many aspects. It was also hypothesized that the model would not be as good a fit for Test area 2 as it was for the Model Study Area. However, the AUC for Test Area 2 (0.846) was significantly higher than that of the Model Study Area (0.725). Looking at AUC to gauge model fitness would inappropriately lead the modeler astray in the assumption that the model is well suited in predicting probability of occurrence in the Test area in this case. It is, as previous studies have suggested, important to use several indices of fitness and to evaluate the model results thoroughly (MacDonald, 2004; Lobo et al. 2008).

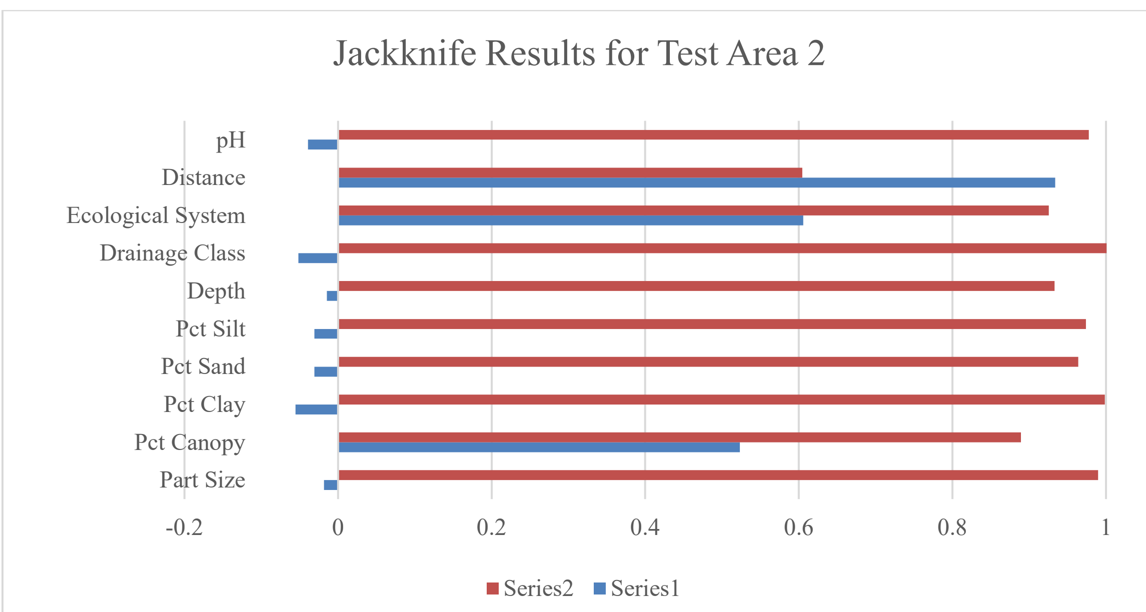


Figure 32: Jackknife of impact to gain by variables included within the model run for Test Area 2.

The Jackknife test for Test Area 2 shows that all but three environmental variables contribute a negative gain when determining the unique information contributed by that layer to the model (blue bar). This can be an indicator of highly correlated data layers but not all of the layers with negative test gain would be correlated. This draws into question the validity of use of this model for Test Area 2 even though the AUC for Test Area 2 was high. PctCanopy contributed useful unique information to the model (blue bar) and has the largest impact, after distance to nearest road, to a reduction in gain when excluded (red bar). Distance to nearest road had the bigest impact on the model as seen in both the jackknife test and the table of variable contributions. The jackknife test shows significant reduction in gain when this variable is excluded from the model. Test Area 2 has significantly more developed area and significantly more mapped road features than the other two areas included in this study. This likely contributes to the significance of road nearness to the resultant model

The three variables that contributed the most significant increase in gain in Test Area 2 when viewed in isolation acording to the jackknife of test gain, were distance to nearest road feature (Distance), ecological system (EcolSys) and percent canopy (PctCanopy) in that order acording to the jackknife. These are the three highest in permutation importance and percent contribution for Test Area 2 as well.

4.4. True Skill Statistic (TSS)

The True Skill Statistic (TSS) is a form of Kappa that is not affected by prevalence or the size of the validation set (Allouche, Tsoar, and Kadmon 2006). Allouche suggests that TSS should be used over Kappa when a threshold-dependent measure is desired (Allouche, Tsoar, and Kadmon 2006). TSS values range from -1 to +1 where values of 0 or less are no better than random and the value of +1 is optimal (Allouche, Tsoar, and Kadmon 2006). TSS, as an indicator of model fitness, was calculated for each replicate run within each study area's model results. The highest TSS value for each model was used as the score for that model in evaluating model fitness using TSS (Table 13). Since TSS is a special case of Kappa and spans the same value range, TSS can be gauged by the same degree of agreement assessment as would Kappa. A value of +1 is perfect agreement, values of 0.75 to 1 represent excellent agreement, 0.4 to 0.75 indicate fair to good agreement and values less than 0.4 are an indication of poor agreement (Monserud and Leemans 1992).

In this study, TSS values for all models were fairly low. TSS for the Model Study Area is considered “Fair” at 0.4087 where TSS for Test Area 1 is just shy of “Fair” at 0.3944. the TSS score for Test Area 2 was “Poor” at 0.2377. Given that the AUC values for the Model Study Area and Test Area 1 were adequate but not stellar (0.725 and 0.746 respectively), a “Fair” TSS value would be expected. For Test Area 2, the TSS is very low given the relatively good AUC (0.846) however when taking the standard deviation of AUC and very low logistic threshold (0.1462) returned from the Test Area 2 model run into account, in this instance TSS helps to confirm that the transfer of the model to Test Area 2 is questionable.

Table 13: TSS for each replicate run for the Model Study Area, Test Area 1 and Test Area 2. The highest TSS of the replicates for each model area was used.

Model	**Replicate 0**	**Replicate 1**	**Replicate 2**	**Replicate 3**	**Replicate 4**
Model Study Area	0.4041	0.4087	0.4064	0.4079	0.4035
Test Area 1	0.3889	0.3938	0.3944	0.3925	0.3918
Test Area 2	0.2377	0.1837	0.1863	0.1401	0.2341

In summary, the Model Study Area Maxent model used in this analysis was evaluated using a 5-fold sub-sample with 50% of the presence points set aside randomly for testing the model. The Model Study Area model was then transferred to Test Area 1 and Test Area 2 and the model suitability results for these three models (Model Study Area Maxent model, Test Area 1 Maxent model, and Test Area 2 Maxent model) were then compared. AUC, test omission rate, TSS, and individual variable contributions were used as indicators of model fitness and transferability success. The results of this study showed acceptable AUC (0.725) and fair TSS (0.409) with a good omission rate (0.0832) for the Model Study Area. For Test Area 1 the AUC was also acceptable (0.746) and TSS fair (0.394) with good omission rate (0.0807). Test Area 2 produced a good AUC (0.8460) but with a poor TSS (0.238) and poorer omission rate than the other models tested (0.2941). Overall the covariates with the most influence on the model, as determined by the permutation importance and review of the Jackknife of test gain, were PctCanopy (56.9) followed by EcolSys (14.7) and soil pH (8.3) for the Model Study Area. Test Area 1's most influential covariates were PctCanopy (30). PctSilt (16.8), and PartSize (16.3). Test Area 2's most influential covariates were Distance (43.5), followed by PctCanopy (36.6), and Bed (7.5).

Chapter 5 Conclusions

The goal of this study was two-fold, to evaluate the fitness for use of Maxent in predicting the potential distribution of cogongrass infestation given suitable conditions within the Model Study Area and to test the transferability of that model to other study areas within the state of Alabama. The guiding objective, beyond the generation of an appropriate model that is transferable across various areas of the state, was the hope that the resulting model and transferability tests would be useful in guiding future survey efforts and funding allocation decisions. In this chapter, we provide a general overview of study concerns as well as results of the study. This chapter concludes by providing a brief narrative on inferences gleaned from the study and potential future work related hereto.

To evaluate the fitness for use of Maxent in predicting the probability of presence distribution of cogongrass within the Model Study Area and to test the transferability of that model to other study areas within the state of Alabama, it was important to thoroughly review the species' biological, climatic and ecological requirements. Cogongrass is highly tolerant to a wide range of conditions and therefore determining the best environmental covariates to use within the model was time consuming. Cogongrass' range of habitat with relation to geographic location (Latitudes 45ºN to 45ºS), rainfall (75 – 500cm average annual), elevation (sea level to 2000m), soil organic matter, habitable sites, and temperature (tolerant to -14ºC) were all generally met within the geographic boundary of the state of Alabama. Review of previous research on the species was used to guide the environmental covariates used in the study, with a focus on land use and soils related variables.

A review of the test areas' output data was performed to determine which environmental factors play the biggest role in transferability hit or miss. The percent contribution and permutation importance of each variable was reviewed along with modeled response curves and the results of the jackknife of regularized test gain. In this study, the most relevant environmental covariate for all three study sites was percent canopy. Percent canopy was the variable with the highest level of permutation importance and percent contribution for both the Model Study Area and Test Area 1 and was the second highest in these factors for Test Area 2. Percent canopy was also in the top three for effect on gain according to the jackknife of regularized test gain graph included in the Maxent output dataset. Therefore, this environmental variable should be included in any future work related to the species. Ecological system, distance to road, percent silt and percent clay also showed significance in this study.

Several of the layers selected for use in the study empirically have some degree of correlation, for instance the particle size layer is a general classification (grouping) of the soil texture as determined by grain size for the topmost horizon of soil using the standards used by the U.S. Department of Agriculture. The individual soil texture layers (percent clay, percent sand, and percent silt) are considered in the particle size layer to some degree. However, since particle size is a classified categorical dataset and the three soil texture layers are discrete measurable values, it was determined that both types of data could be included without issue. Correlation between datasets should be taken into consideration when analyzing SDMs such as that produced in this study.

5.1. Uncertainty in the Model

This model may be used to support decisions related to where to survey for cogongrass locations and what counties to focus on for eradication efforts. Therefore, it is important that the uncertainty in the model be clearly understood so that the value of the model results can be articulated to stakeholders. Sample selection bias is a fundamental limitation of presence only modeling such as is the case in this study using Maxent. This bias can have a significant impact on the model outcome (Elith 2011; Phillips et al. 2009). Examples of sample selection bias can be found in this study and in the Ervin and Holly (2011) study, which specifically focused on a biased sample by sampling along roadways.

In the current study, sample selection bias is introduced by the method of discovery and subsequent reporting of suspected cogongrass location points to the AFC. The AFC relies heavily on landowner and public reporting of suspected point locations and then investigates and verifies those locations. This bias cannot be removed due to the nature of infestation reporting; however, it is prudent to take it into account when analyzing the model result. Also, some sampling bias can be removed from the study in areas where the species presence point data is not uniformly scattered across the entire extent of the test area's geography space by use of a minimum bounding geometry layer that will ensure that any test or background pseudo-presence data predictions will use the same geographic boundary as the training data.

5.2. Proposed Future Work

Future work related to this study should include the testing of transferability across additional AFC Work Units and potentially recalibrating and retesting models as new data becomes available (Stohlgren and Schnase 2006). It would be appropriate to test the model against all AFC Work Units in a future study as differences in model performance was noted between the areas included in this study. Further, in the Stohlgren and Schnase study, it was recommended that the modeling process be an iterative process in which the model is recalibrated as new species presence data become available (Stohlgren and Schnase 2006; Crall et al. 2013). Thus, if a model is used to prompt a guided species survey, the survey results can then be added to the volume of existing species presence point data and the model can be rerun to create a new model with this updated sample layer to better inform the next guided survey. This would be especially important if the output of the model was to be used to guide funding for control and eradication efforts in the future.

Additionally, field verification of model output to determine if the predicted locations do, in fact, support cogongrass infestations would be useful. And finally, further work into transportation corridor related factors on the distribution of cogongrass should also be considered. Specifically, the roads layer used in this study did not contain all interior local roads, especially in heavily timbered and rural locations. Since Alabama has a high percentage of forested area, it would be prudent to launch a project to update the roads layer used in the distance to roads calculation or pursue the purchase or construction of a better suited roads layer.

5.3. Findings

Given the acceptable AUC, omission rate and TSS values of the original Model Study Area's Maxent model output, the model produced for this study can be considered to be an appropriate model for predicting the presence of cogongrass in the Model Study Area. The transferability test for the model leaves some open questions, however. The results of this study showed that when the area targeted for transfer is similar environmentally and geographically to the Modeled Study Area, this model can perform sufficiently well to be used to inform the analyst on predicted probability in the target area. When the target area is highly dissimilar, as is the case with Test Area 2 in this study, caution should be taken when transferring the model to this new geographic space. It would perhaps be more valid to re-evaluate the model against the new geographic area and re-run with a modified set of covariates as appropriate.

In summary, the model produced by Maxent for the Model Study Area had an AUC of 0.725 which is considered to be acceptable for use in conservation planning (Elith et al. 2006).

The environmental covariates selected for the study were suitably broad in their biological and ecological suitability to the species being studied to allow for successful transfer of the model to two other AFC Work Units within the state, however detailed review of the model results using multiple metrics for testing fitness should be employed when verifying model transferability success.

This study adds to the body of work related to species distribution modeling using Maxent for cogongrass as well as transferability studies of Maxent models for invasive species in general. Although additional work is suggested to further this study of transferability of Maxent model for cogongrass, the findings of this study suggest that Maxent is potentially a suitable tool for modeling the predicted potential distribution of cogongrass infestation given suitable biological and ecological variables are utilized. This study also suggests that a suitably trained Maxent model can be successfully projected to similar geographic areas within a limited extent, such as a state as was tested here. The transfer of a suitably trained Maxent model to an area of dissimilar geographic or environmental conditions, should be accepted with caution.

References

Akobundu, I. O., and F. E. Ekeleme. 2000. "Effect of method of *Imperata cylindrica* management on maize grain yield in the derived savanna of south-western Nigeria." *Weed Research*. 40(4): 335–341.

Alabama Forestry Commission. n.d. "Cogongrass." Accessed April 18th, 2018. http://www.forestry.state.al.us/Pages/Informational/Invasive/Cogongrass.aspx.

Alabama Forestry Commission. n.d. "Cogongrass viewer." Accessed April 18th, 2018. http://www.forestry.state.al.us/viewers/afc_cogongrass_viewer.aspx

Allouche, Omri, Asaf Tsoar, and Ronen Kadmon. 2006. "Assessing the accuracy of species distribution models: prevalence, Kappa and the true skill statistic (TSS)." *Journal of Applied Ecology* 43 (6):1223-1232. doi: 10.1111/j.1365-2664.2006.01214.x.

Anderson, R.P. 2012. "Harnessing the world's biodiversity data: promise and peril in ecological niche modeling of species distributions." *Annals of the New York Academy of Sciences* 1260: 66-80.

Ayeni, A.0. 1985. "Observations on the vegetative growth pattern of speargrass (*Imperata cylindrica* (L.) Beauv.)." *Agriculture. Ecosystems and Environment* 13 (3-4): 301-307. doi: 10.1016/0167-8809(85)90017-9.

Clemson University. N.d. "Clemson Regulatory Services." Accessed November 20, 2018. https://www.clemson.edu/public/regulatory/plant-protection/invasive/cogongrass/index.html

Coulston, J. W., Gretchen G. Moisen, B.T. Wilson, M.V. Finco, W.B. Cohen, and C.K. Brewer, 2012. "Modeling percent tree canopy cover: a pilot study." *Photogrammetric Engineering & Remote Sensing* 78(7): 715-727.

Crall, Alycia W., Catherine S. Jarnevich, Brendon Panke, Nick Young, Mark Renz, and Jeffrey Morisette. 2013. "Using habitat suitability models to target invasive plant species surveys." *Ecological Applications* 23: 60-72. doi: 10.1890/12-0465.1.

Databasin. September 3, 2014. "GAP Land Cover Data for Alabama, USA" Last Accessed March 14, 2019. https://databasin.org/datasets/e6c2c82715be44bba3579fa6010acfd5.

Dozier, Hallie. Sandra K. Gaffney, Eric McDonald, R.R.L. Johnson, and Donn G. Shilling. 1998. "Cogongrass in the United States: History, Ecology, Impacts, and Management." *Weed Technology*. 12: 737-743.

Dickens, R. 1974. "Cogongrass in Alabama after sixty years." *Weed Science* 22: 177-179.

EDDMapS. 2019. “Early Detection & Distribution Mapping System.” Accessed August 4, 2019. The University of Georgia - Center for Invasive Species and Ecosystem Health. http://www.eddmaps.org/.

Elith, Jane, C.H. Graham, Robert P. Anderson, Miroslav Dudík, Simon Ferrier, Antoine Guisan, Robert J. Hijmans, et al. 2006. “Novel Methods Improve Prediction of Species’ Distributions from Occurrence Data.” *Ecography* 29, no. 2(April): 129-151.

Elith, Jane, Steven J. Phillips, Trevor Hastie, Miroslav Dudik, Yung En Chee, and Colin J. Yates. 2011. “A statistical explanation of Maxent for ecologists.” *Diversity and Distributions* 17: 43-57.

Enloe, S.F., D.K. Lauer, N.J. Loewenstein, and R.D. Lucardi. 2018. “Response of twelve Florida cogongrass (*Imperata cylindrica*) populations to herbicide treatment.” *Invasive Plant Science and Management* 11(2): 82-88.

Ervin, Gary N., and D.C. Holly. 2011. “Examining Local Transferability of Predictive Species Distribution Models for Invasive Plants: An Example with Cogongrass (*Imperata cylindrica*).” *Invasive Plant Science and Management* 4 (4): 390-401.

Esri. n.d. “ArcMap: Extract by Mask” Accessed August 1st, 2019. http://desktop.arcgis.com/en/arcmap/10.6/tools/spatial-analyst-toolbox/extract-by-mask.htm.

Estrada, James A., and S. Luke Flory. 2014. “Cogongrass (*Imperata cylindrica*) invasions in the US: Mechanisms, impacts, and threats to biodiversity.” *Global Ecology and Conservation* 3: 1-10.

Eussen, J.H.H., and S. Wirjahardja. 1973. “Studies of an alang-alang, *Imperata cylindrica* (L.) Beauv. vegetation.” *Biotropica Bullitan*. no. 6.

Gaffney, J.F. 1996. “Ecophysiological and Technical Factors Influencing the Management of Cogongrass (*Imperata cylindrica*).” Ph.D. dissertation, University of Florida.

Halvorsen, Rune, Sabrina Mazzoni, John Wirkola Dirksen, Erik Næsset, Terje Gobakken, and Mikael Ohlson. 2016. “How Important Are Choice of Model Selection Method and Spatial Autocorrelation of Presence Data for Distribution Modelling by Maxent?” *Ecological Modelling* 328: 108–118.

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

Holm, L.G., D.L. Pucknett, J.B. Pancho, and J.P. Herberger. 1977. “The World’s Worst Weeds. Distribution and Biology”. Honolulu HI: University of Hawaii Press.

Howard, Janet L. 2005. “*Imperata brasiliensis, I. cylindrica*. In: Fire Effects Information System.” Accessed March 24th, 2019. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory. https://www.fs.fed.us /database/feis/plants/graminoid/impspp/all.html.

Hubbard, C.E. 1944. *"Imperata cylindrica*. Taxonomy, Distribution, Economic significance, and Control." *Agricultural Bureau Joint Publication*. No. 7, Imperial Bureau Pastures and Forage Crops, Aberystwyth, Wales. Great Britton.

King, Sharon E., and James B. Grace, 2000. "The Effects of Soil Flooding on the Establishment of Cogongrass (*Imperata cylindrica*), a Nonindigenous Invader of the Southeastern United States." *Wetlands* 20(2): 300-306.

Lee, S.A. 1977. "Germination, rhizome survival, and control of *Imperata cylindrica*(L.) Beauv. on peat." *MARDI Research. Bulletin* 5(2): 1-9.

Lippincott, Carol L. 1997. "Ecological consequences of *Imperata cylindrica* (cogongrass) invasion in Florida sandhill." Dissertation, University of Florida.

Lippincott, C.L. 2000. "Effects of *I. cylindrica* (cogongrass) invasions on fire regimes in Florida sandhill." *Natural Area Journal* 20: 140–149.

Livingston, M.J., and C. Osteen. 2008. "Integrating Invasive Species Prevention and Control Policies." *Economic Brief No. 11.* USDA, Economic Research Service.

Lucardi, Rima, Lisa Wallace, and Gary Ervin. 2014. "Invasion Success in Cogongrass (*Imperata cylindrica*): A Population Genetic Approach Exploring Genetic Diversity and Historical Introductions." *Invasive Plant Science and Management* 7, no. 1: 59–75.

MacDonald, G.E. 2004. "Cogongrass (*Imperata cylindrica*)-Biology, Ecology, and Management." *Critical Reviews in Plant Sciences*. 23(5): 367-380.

McNeely, Jeffrey A. (ed.). 2001. *The Great Reshuffling: Human dimensions of invasive alien species*. IUCN, Gland, Switzerland and Cambridge, UK.

Merow, Cory, Matthew J. Smith, and John A. Silander. 2013. "A Practical Guide to Maxent for Modeling Species' distributions: What it Does, and Why Inputs and Settings Matter." *Echography* 36(10): 1058-1069. doi: 10.1111/j.1600-0587.2013.07872.x.

Monserud, Robert A., and Rik Leemans. 1992. "Comparing global vegetation maps with the Kappa statistic." *Ecological Modelling*, 62: 275-293.

Narkhede, Sarang. 2018. Understanding AUC-ROC Curve. Towards Data Science. Posted June 26th, 2018. Accessed 11/5/2018. https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5

O'Sullivan, David, and George W. Perry. 2013. *Spatial Simulation: Exploring Pattern and Process*. London and New York: John Wiley & Sons.

Patterson, D.T. 1980. “Shading effects on growth and partitioning of plant biomass in cogongrass (*Imperata cylindrica*) from shaded and exposed habitats.” *Weed Science.* 28: 735-740.

Peterson, A.T., J. Soberon, R.G. Pearson, R.P. Anderson, E. Martınez-Meyer, M. Nakamura, and M.B. Araujo. 2011. *Ecological niches and geographic distributions*. Princeton University Press, Princeton, NJ.

Phillips, Steven J. 2017. “A Brief Tutorial on Maxent.” Accessed 1/5/2019. http://biodiversityinformatics.amnh.org/open_source/Maxent/.

Phillips, Steven J., and Miroslav Dudik. 2007. “Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation.” *Ecography* 31: 161-175

Phillips, Steven J., Miroslav Dudík, and Robert E. Schapire, “Maxent software for modeling species niches and distributions” (Version 3.4.1). Accessed on 1/5/2019. http://biodiversityinformatics.amnh.org/open_source/Maxent.

Radosavljevic, Aleksandar and Robert P. Anderson. 2014. “Making better MAXENT models of species distributions: complexity, overfitting and evaluation.” *Journal of Biogeography* 41: 629-643.

Rauschert, Emily S.J., David A. Mortensen, and Steven M. Bloser. 2017. “Human-mediated dispersal via rural road maintenance can move invasive propagules.” *Biological Invasions* 19: 2047-2058. doi: 10.1007/s10530-017-1416-2.

Sajise, P.E. 1976. “Evaluation of cogon (*Imperata cylindrica*) as a serial stage in Philippine vegetational succession. 1. The cogonal seral stage and plant succession. 2. Autecological studies on cogon. *Dissertation Abstracts International B*: 3040-3041. *Weed Abstracts* no. 1339.

Stohlgren, T.J., and J.L. Schnase. 2006. “Risk analysis for biological hazards: What we need to know about invasive species.” *Risk Analysis* 26: 163-173.

Tabor, P. 1949. “Cogongrass, *Imperata cylindrica* (L.) Beauv., in the southeastern United States.” *Agronomic Journal*. 41: 270.

Tabor, P. 1952. “Comments on cogon and torpedograsses: A challenge to weed workers.” *Weeds*. 1: 374-375.

Terry, P.J., G. Adjiers, I.O. Akobundu, A.U. Anoka, M.E. Drilling, S. Tjitrosemito, and M. Utomo. 1997. “Herbicides and mechanical control of *Imperata cylindrica* as a first step in grassland rehabilitation.” *Agroforestry Systems*. 36:151–179

West, Amanda M., Sunil Kumar, Cynthia S. Brown, Thomas J. Stohlgren, and Jim Bromberg. 2016. “Field validation of an invasive species Maxent model.” *Ecological Informatics*. 36: 126-134. doi:10.1016/j.ecoinf.2016.11.001

Wilcut, J.W., R.D. Dute, B. Truelove, and D.E. Davis. 1988a. "Factors limiting the distribution of cogongrass, *Imperata cylindrica*, and torpedograss, *Panicum repens*." *Weed Science*. 36: 577-582.

Wilcut, J.W., B. Truelove, D.E. Davis, and J.C. Williams. 1988b. "Temperature factors limiting the spread of cogongrass (*Imperata cylindrica*) and torpedograss (*Panicum repens*)." *Weed Science.* 36: 49-55.

Willard, T.R., D.W. Hall, D.G. Shilling, J.A. Lewis, and W.L. Currey. 1990. "Cogongrass (*Imperata cylindrica*) distribution on Florida highway rights-of-way." *Weed Technology*. 4: 658-660.

U.S. Department of Agriculture. n.d. "SSURGO Soil Map Coverage versus the U.S. General Soil Map Coverage." Accessed April 10, 2019. https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/geo/?cid=nrcs142p2_053626

U.S. Department of Agriculture. n.d. "NRCS Plants Database". Accessed April 10, 2019. https://plants.usda.gov/core/profile?symbol=IMCY

U.S. Fish and Wildlife Service. n.d. "Invasive Species." www.fws.gov/invasives/faq.html#q2

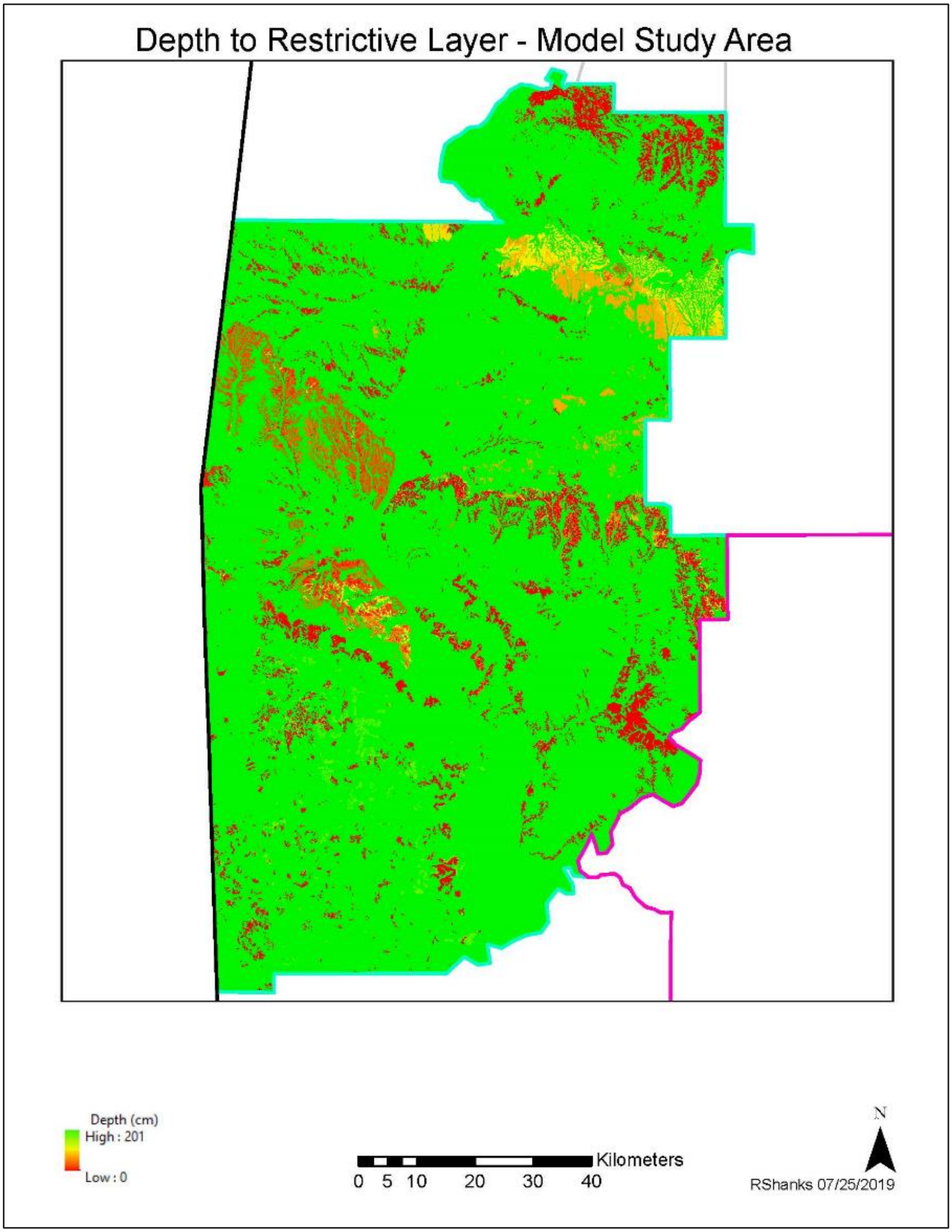
U.S. Forest Service. n.d. "Invasive Species Profile." https://www.fs.fed.us/invasivespecies/speciesprofiles/documents/cogon-grass.pdf

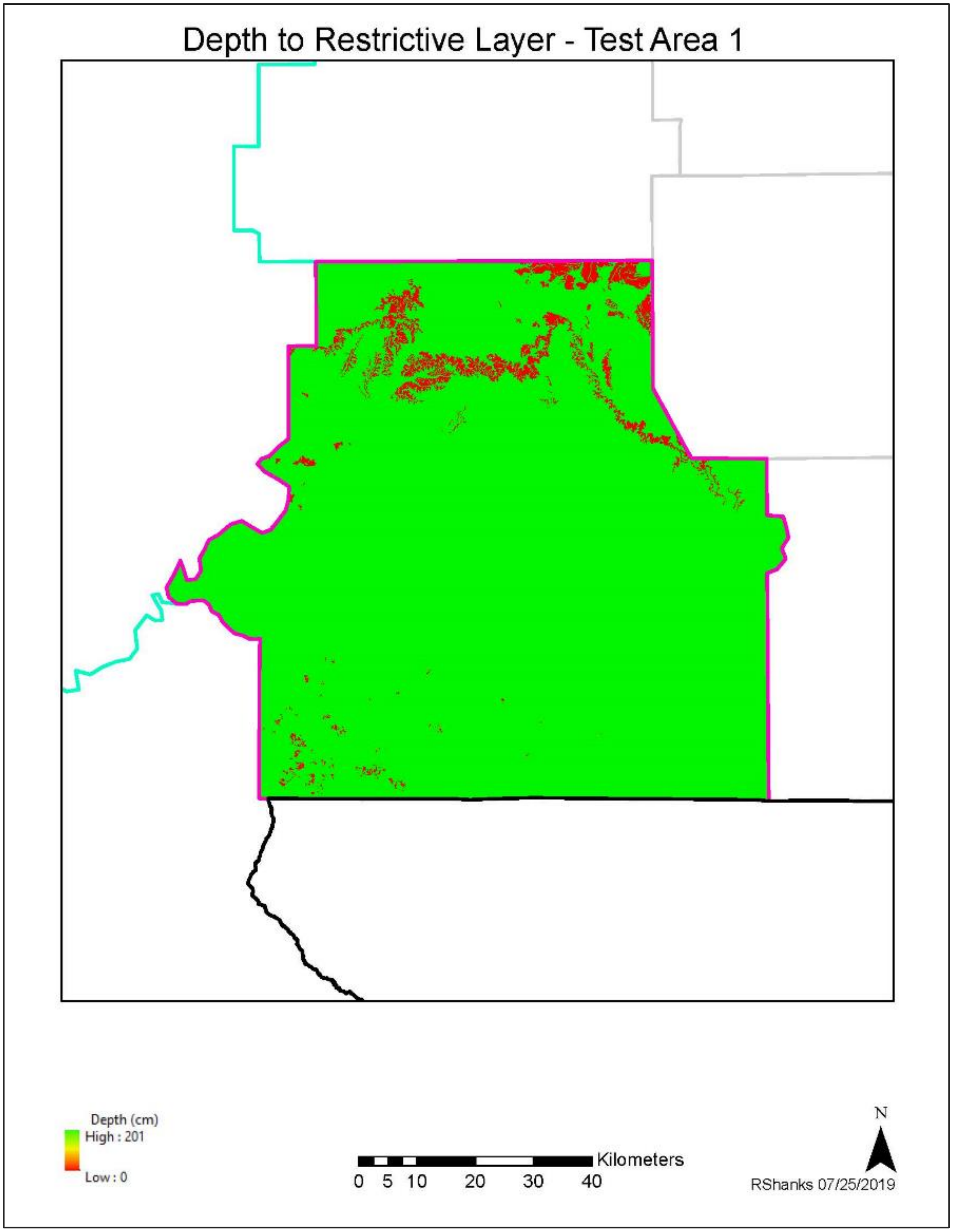
Yager, Lisa, Deborah Miller, and Jeanne Jones. 2011. "Woody Shrubs as a Barrier to Invasion by Cogongrass (*Imperata cylindrica*)" *Invasive Plant Science and Management* Apr-Jun 2011, Vol.4(2), pp. 207-211. Doi: 10.1614/IPSM-D-10-00052.1.

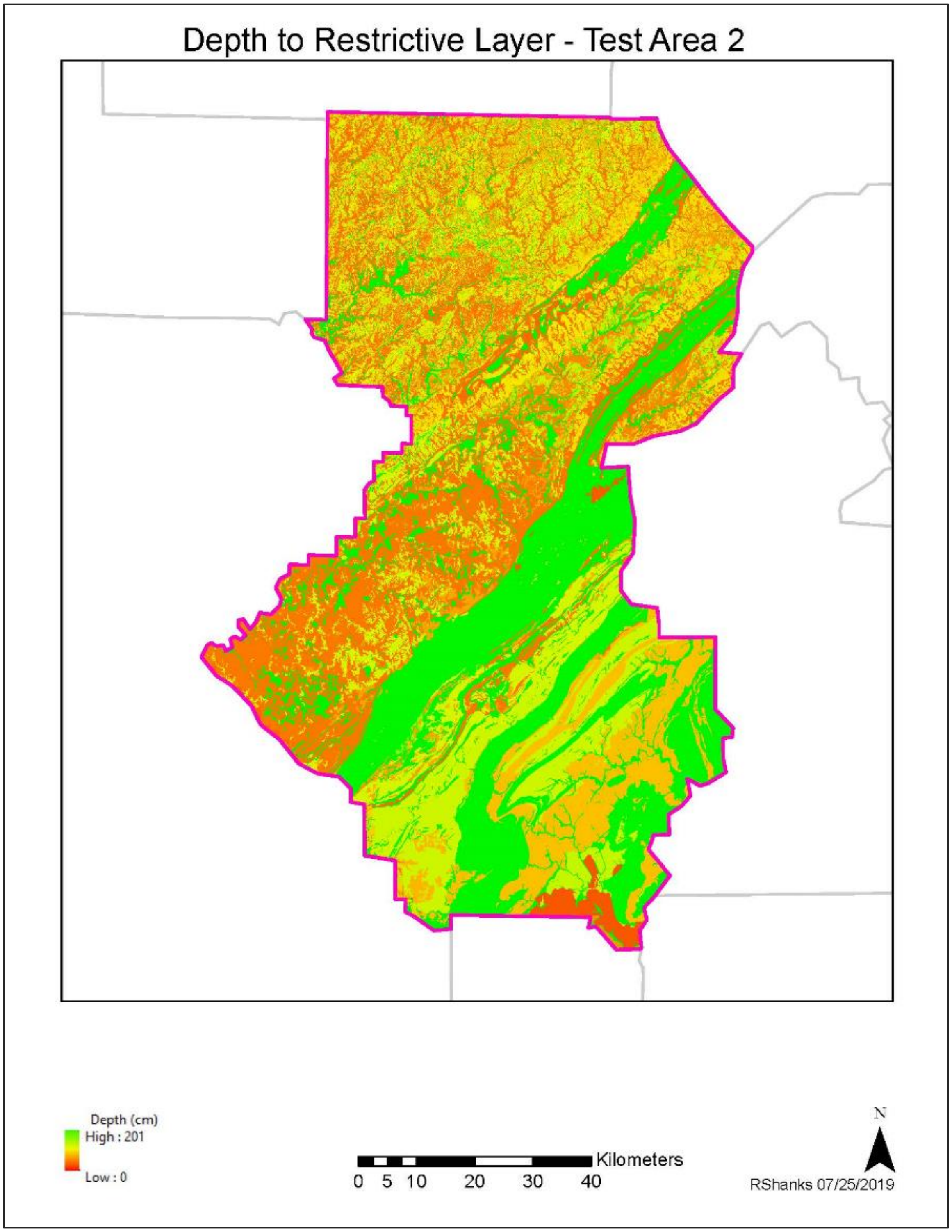
Young, Nick, Carter Lane, and Paul Evangelista. 2011. "A Maxent Model v3.3.3e Tutorial (ArcGISv10)." Natural Resource Ecology Laboratory at Colorado State University and the National Institute of Invasive Species Science.

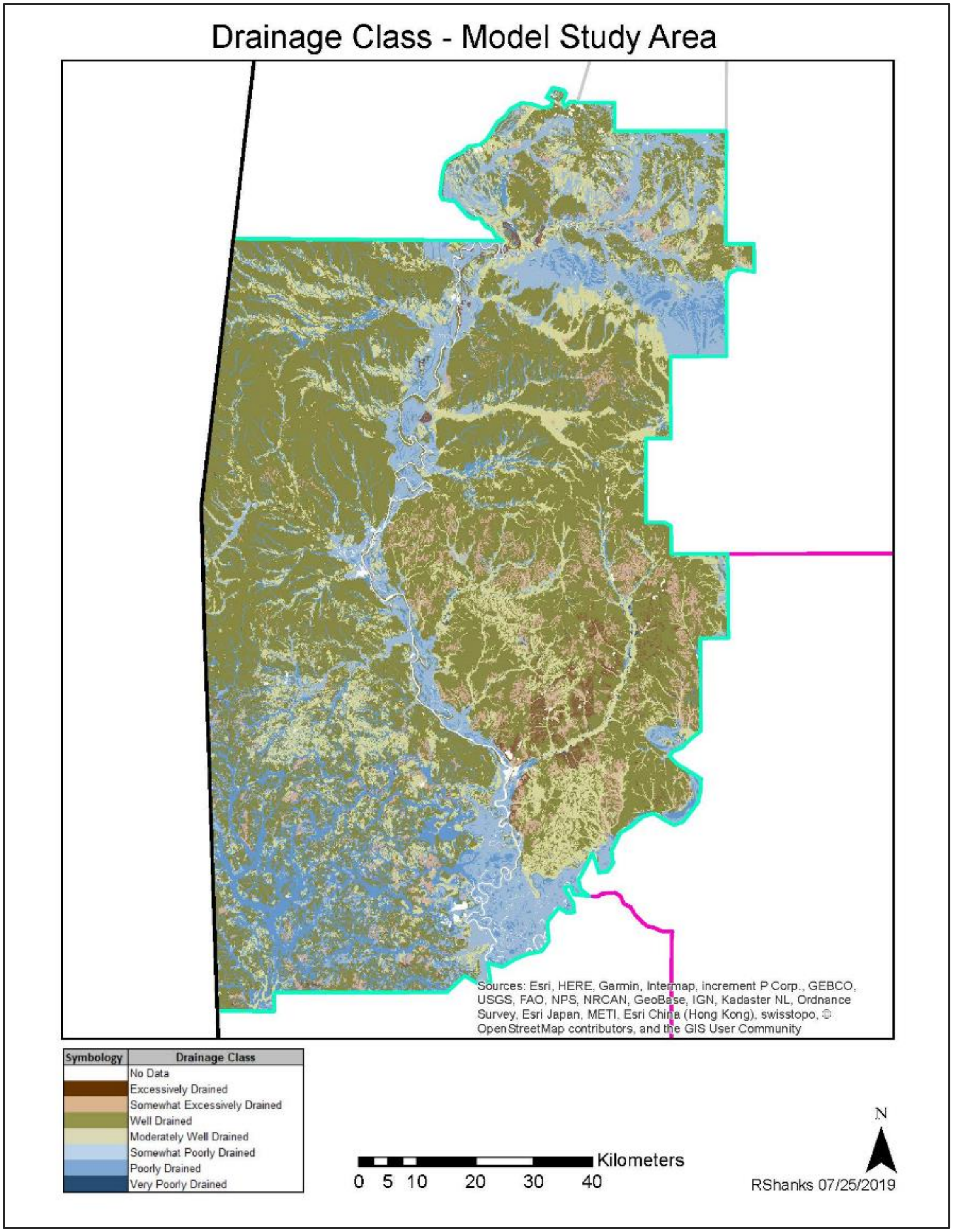
Yu, Feng. 2013. "Improving Model Performance For Invasive Plant Species Distribution Using Global-Scale Presence-Only Data: Parameterization And Data Quality". Master of Science Thesis. Purdue University. https://docs.lib.purdue.edu/open_access_theses/112

Appendix A: Soils Related Environmental Covariate Maps

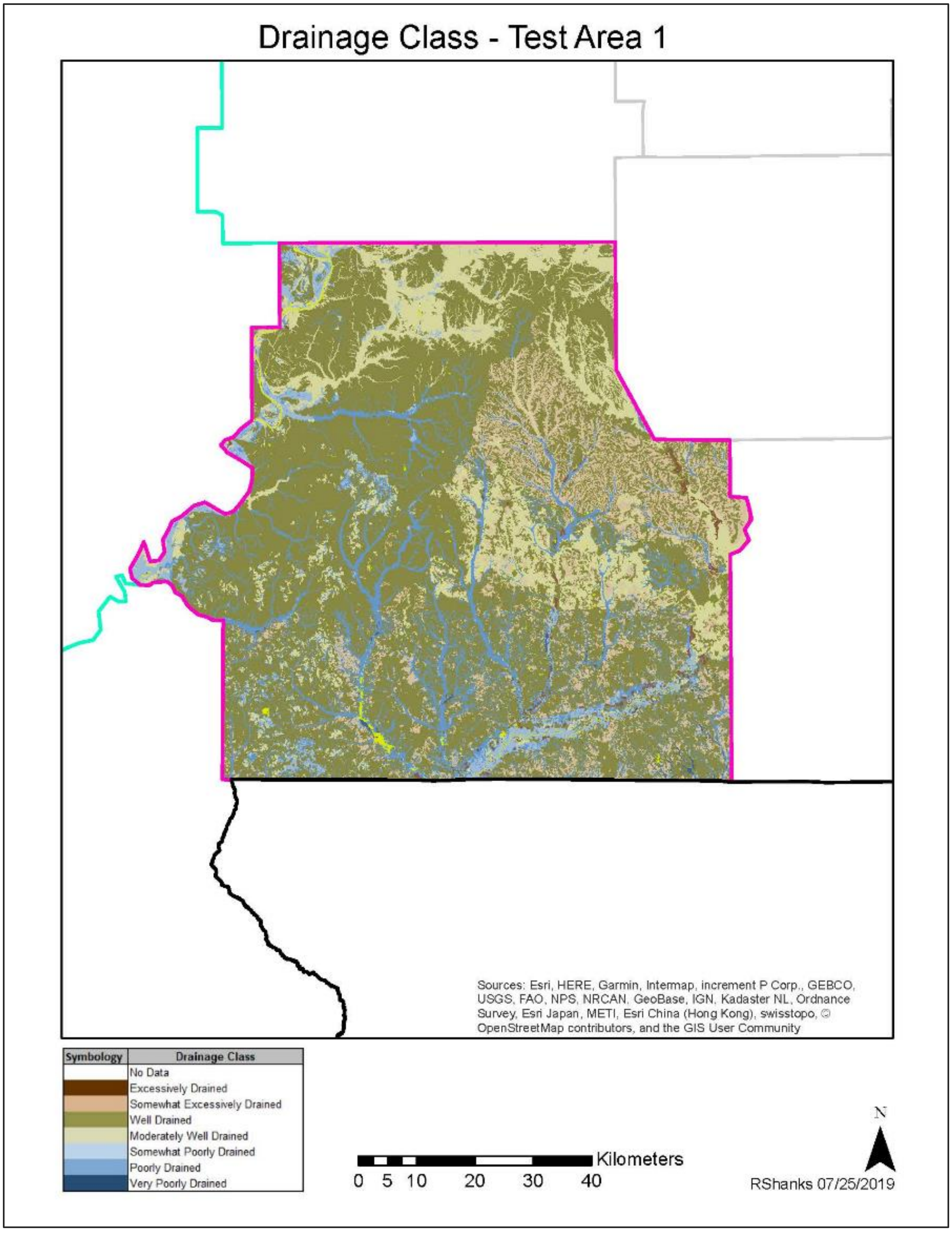



Depth to Restrictive Layer - Test Area 1
Depth (cm)
High : 201
Low : 0
Kilometers
0 5 10 20 30 40
N
RShanks 07/25/2019

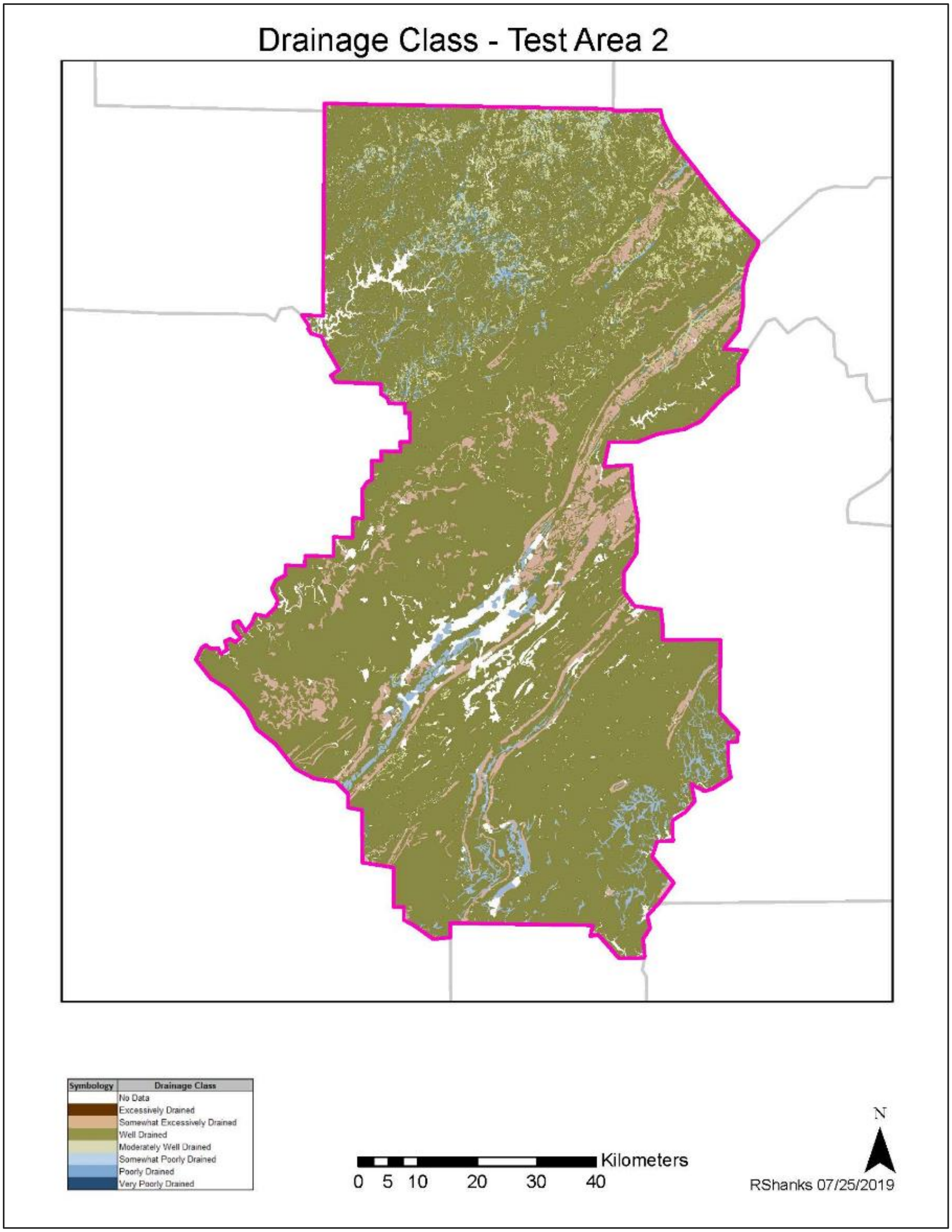

Depth to Restrictive Layer - Test Area 2
Depth (cm)
High : 201
Low : 0
Kilometers
0 5 10 20 30 40
N
RShanks 07/25/2019



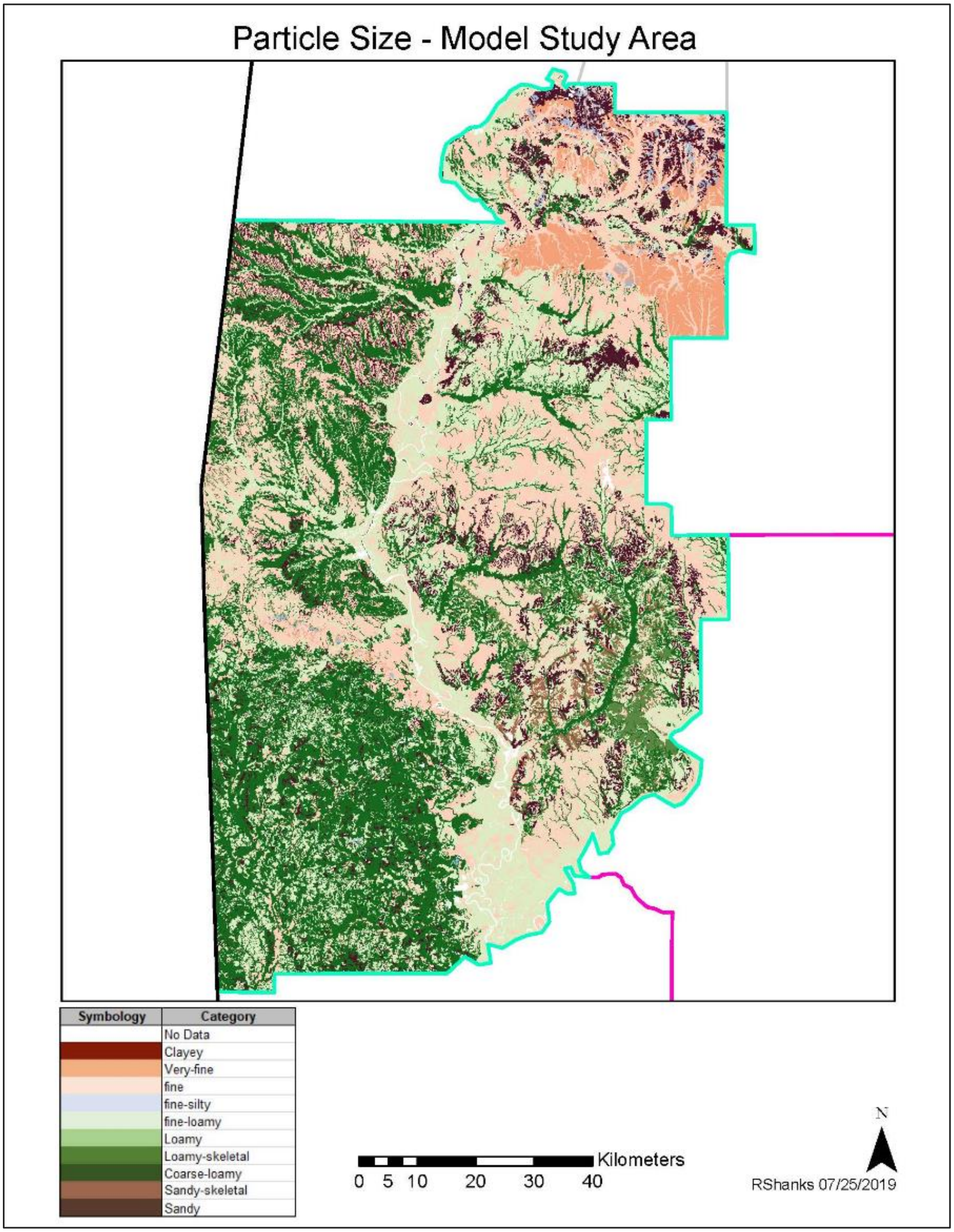
Drainage Class - Model Study Area
Sources: Esri, HERE, Garmin, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, © OpenStreetMap contributors, and the GIS User Community
Symbology
Drainage Class
No Data
Excessively Drained
Somewhat Excessively Drained
Well Drained
Moderately Well Drained
Somewhat Poorly Drained
Poorly Drained
Very Poorly Drained
0 5 10 20 30 40
Kilometers
N
RShanks 07/25/2019

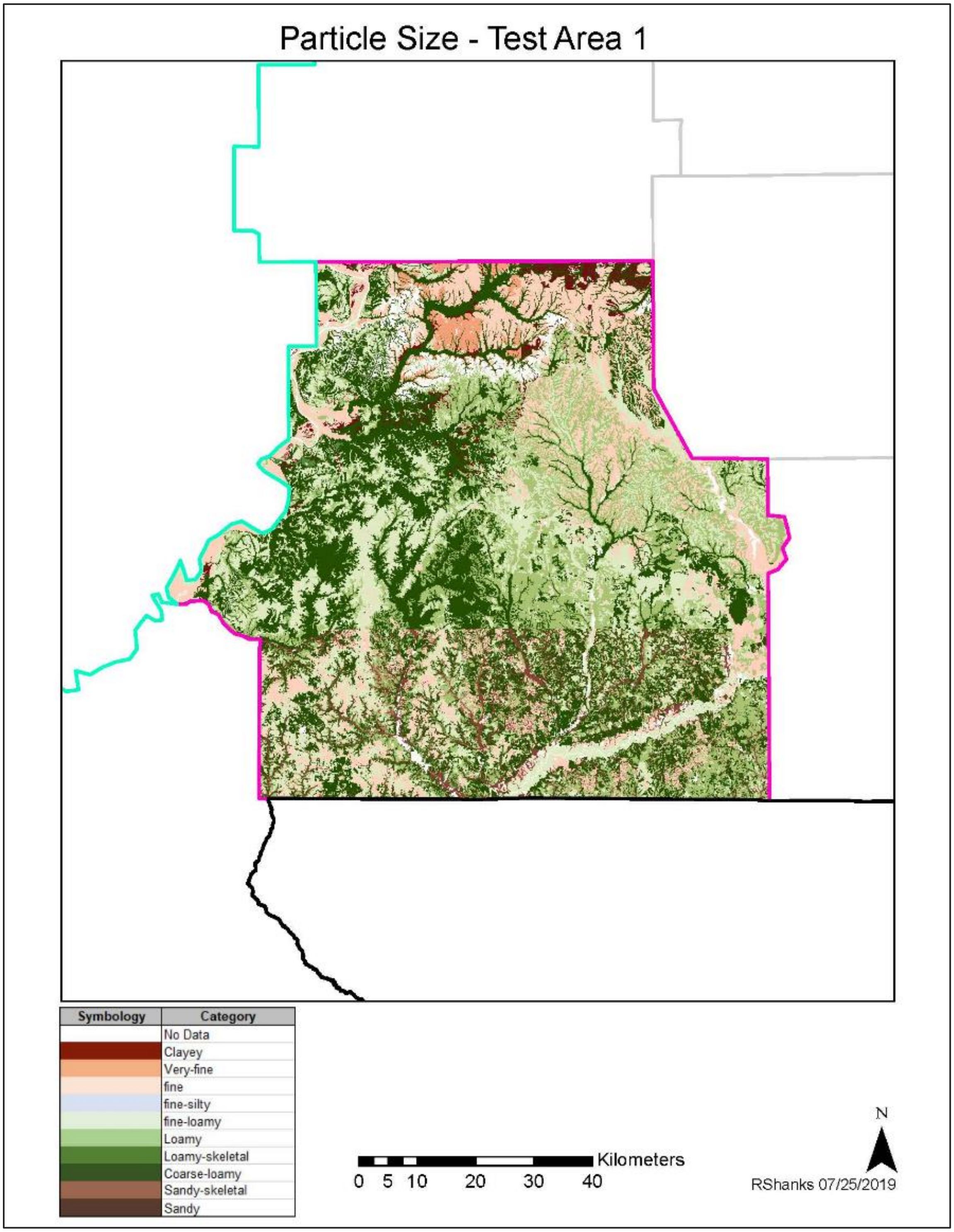


Drainage Class - Test Area 1
Sources: Esri, HERE, Garmin, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, © OpenStreetMap contributors, and the GIS User Community
Symbology
Drainage Class
No Data
Excessively Drained
Somewhat Excessively Drained
Well Drained
Moderately Well Drained
Somewhat Poorly Drained
Poorly Drained
Very Poorly Drained
0 5 10 20 30 40
Kilometers
N
RShanks 07/25/2019

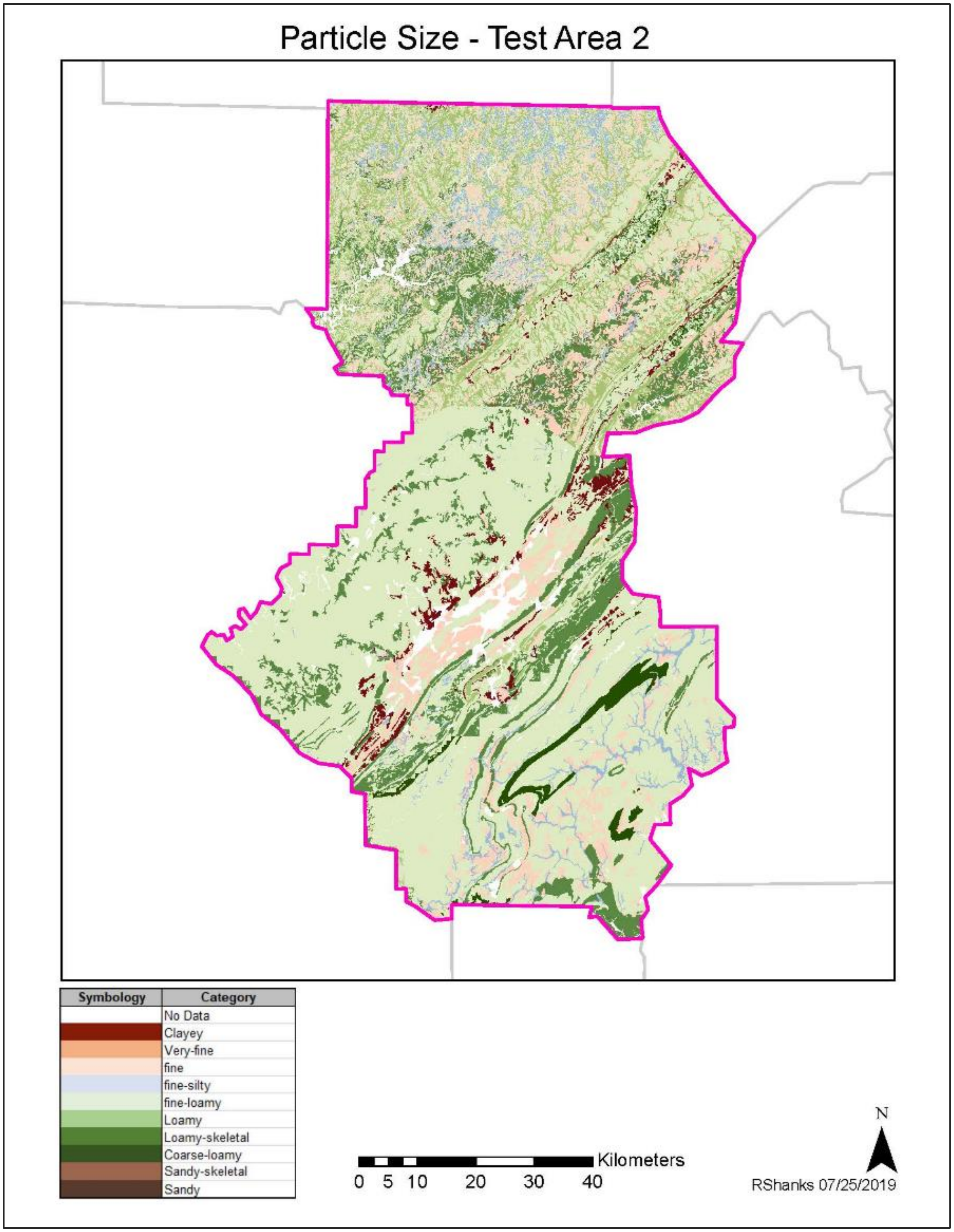


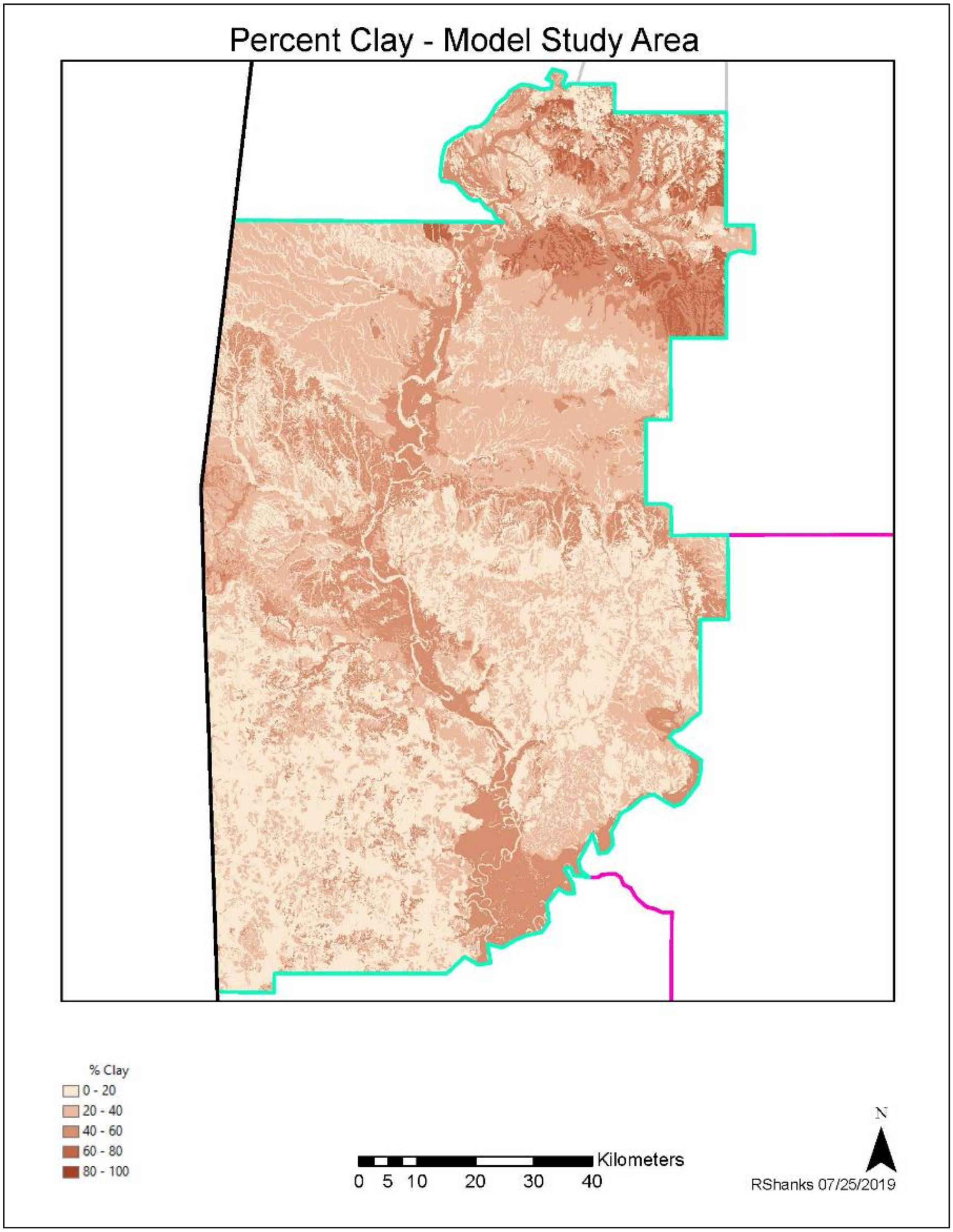
Drainage Class - Test Area 2
Symbology
Drainage Class
No Data
Excessively Drained
Somewhat Excessively Drained
Well Drained
Moderately Well Drained
Somewhat Poorly Drained
Poorly Drained
Very Poorly Drained
Kilometers
0 5 10 20 30 40
N
RShanks 07/25/2019

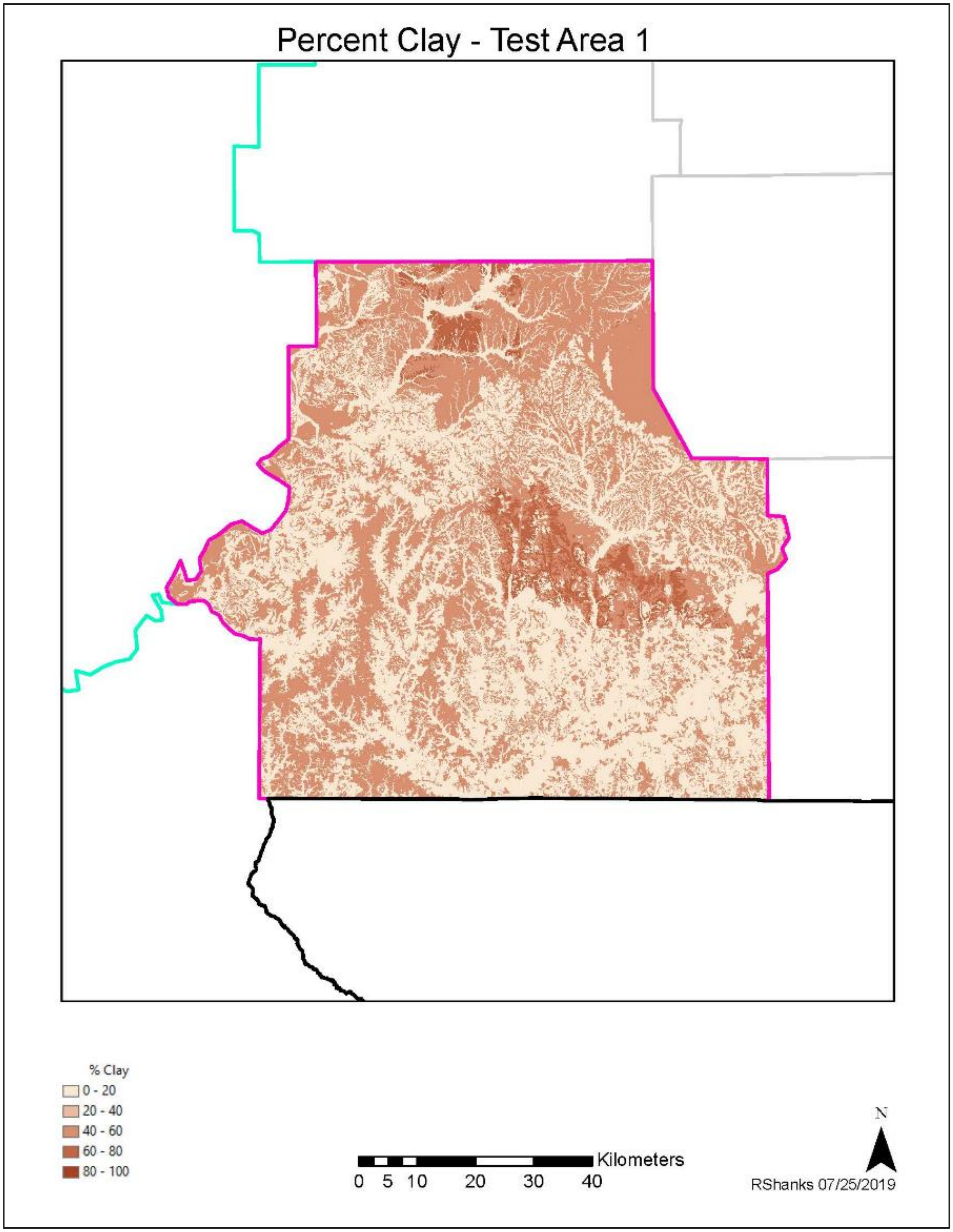

Particle Size - Model Study Area
Symbology
Category
No Data
Clayey
Very-fine
fine
fine-silty
fine-loamy
Loamy
Loamy-skeletal
Coarse-loamy
Sandy-skeletal
Sandy
Kilometers
0 5 10 20 30 40
N
RShanks 07/25/2019

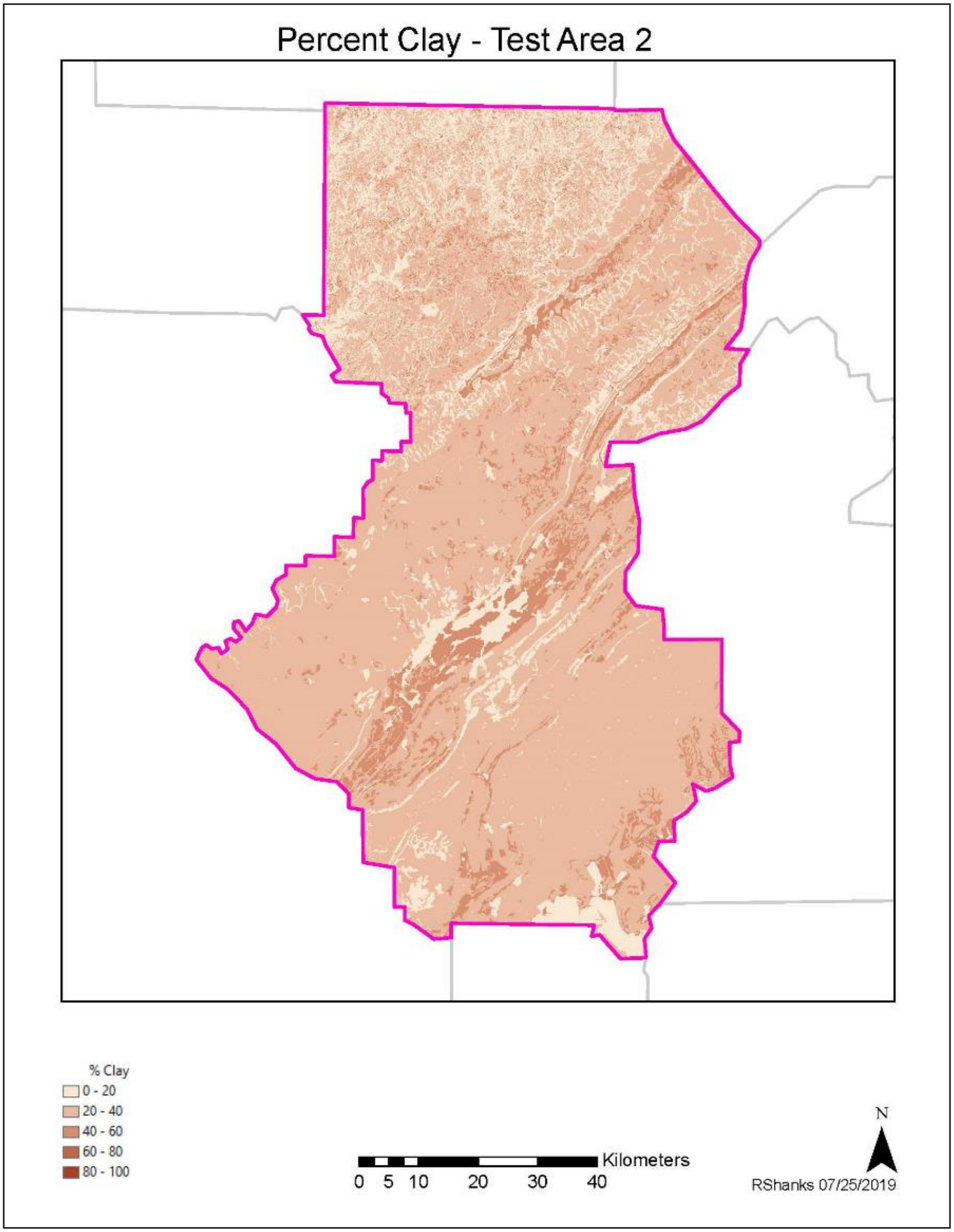


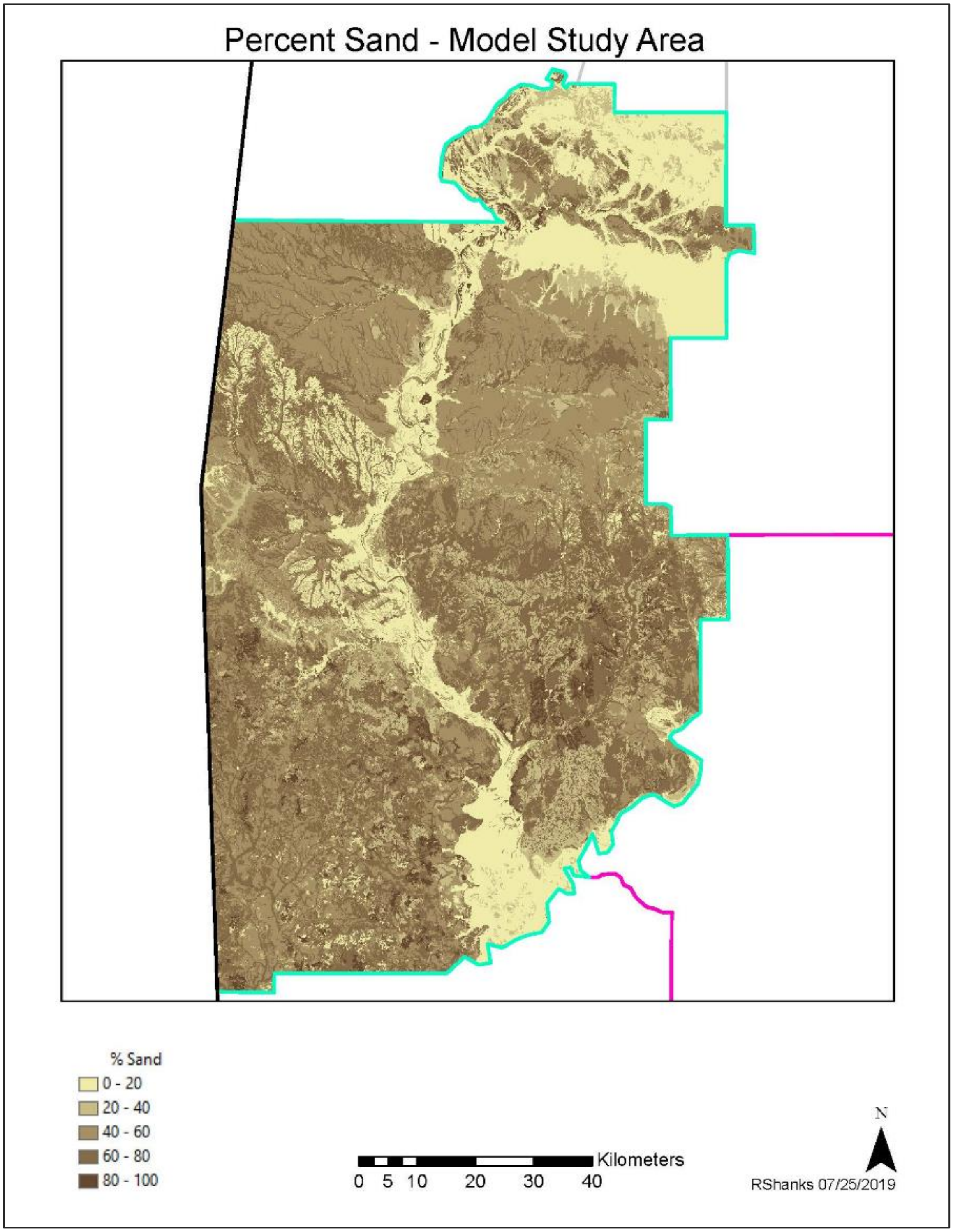
Particle Size - Test Area 1
Symbology
Category
No Data
Clayey
Very-fine
fine
fine-silty
fine-loamy
Loamy
Loamy-skeletal
Coarse-loamy
Sandy-skeletal
Sandy
Kilometers
0 5 10 20 30 40
N
RShanks 07/25/2019

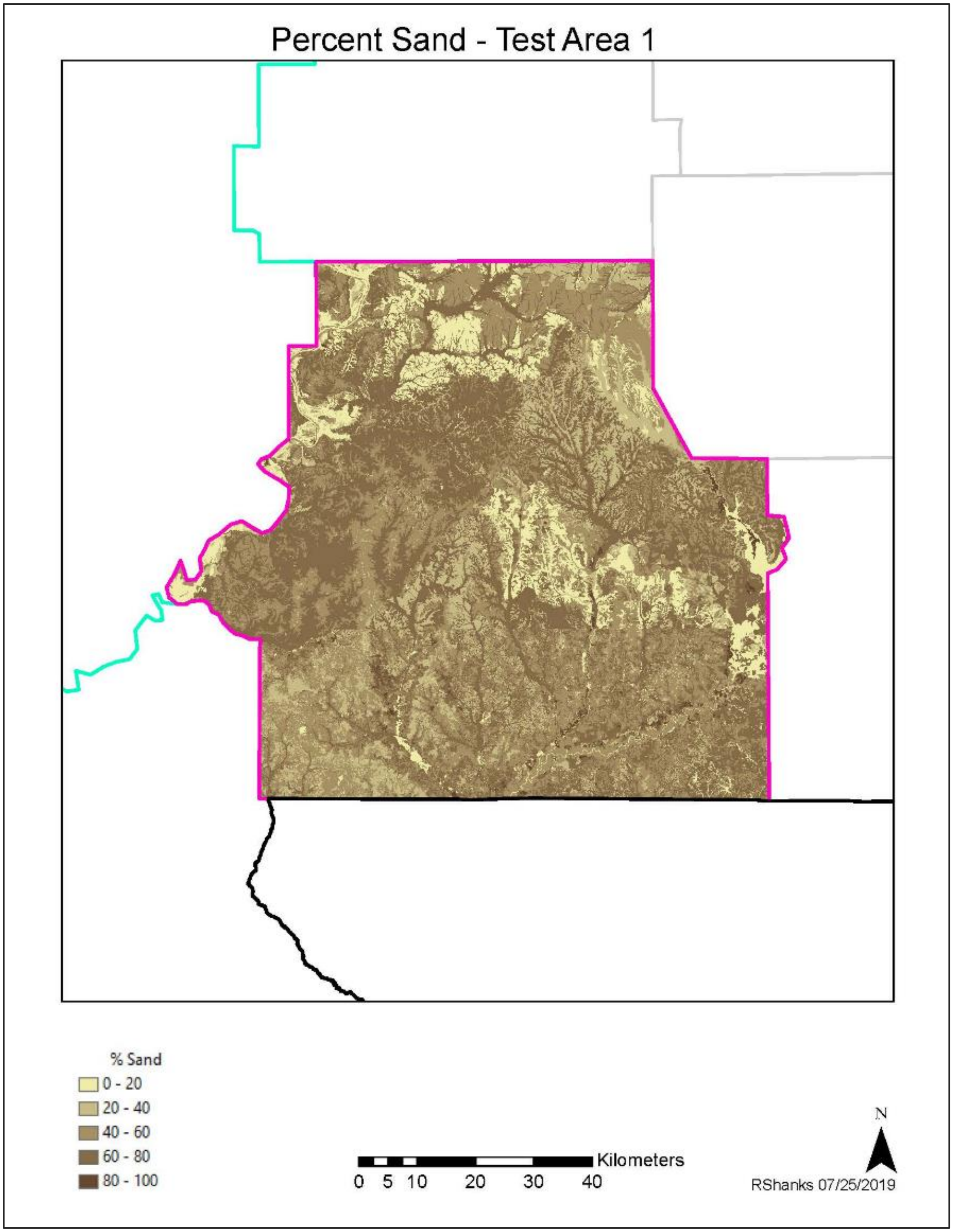

Particle Size - Test Area 2
Symbology
Category
No Data
Clayey
Very-fine
fine
fine-silty
fine-loamy
Loamy
Loamy-skeletal
Coarse-loamy
Sandy-skeletal
Sandy
Kilometers
0 5 10 20 30 40
N
RShanks 07/25/2019

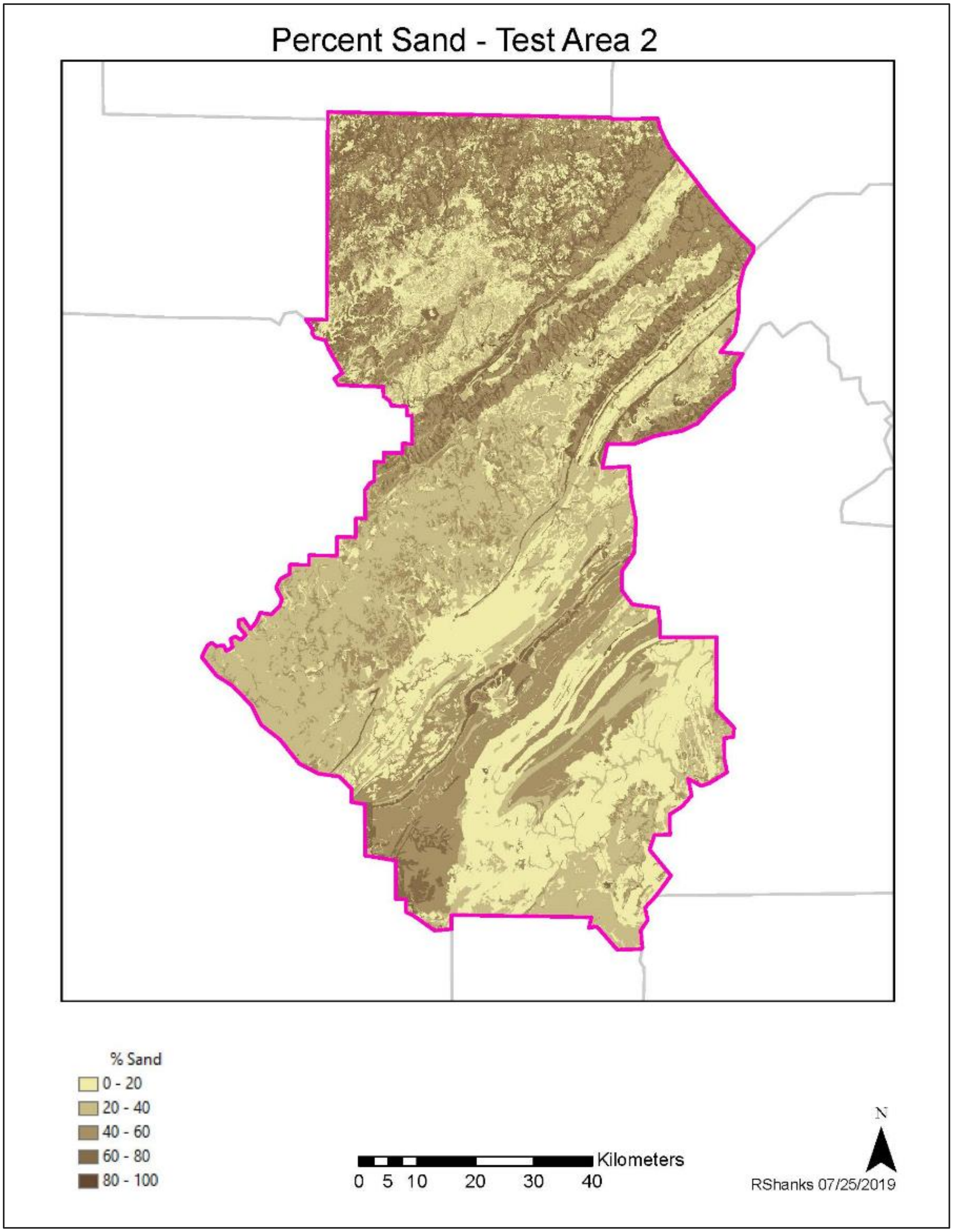

Percent Clay - Model Study Area
% Clay
0 - 20
20 - 40
40 - 60
60 - 80
80 - 100
0 5 10 20 30 40
Kilometers
N
RShanks 07/25/2019


Percent Clay - Test Area 1
% Clay
0 - 20
20 - 40
40 - 60
60 - 80
80 - 100
0 5 10 20 30 40
Kilometers
N
RShanks 07/25/2019

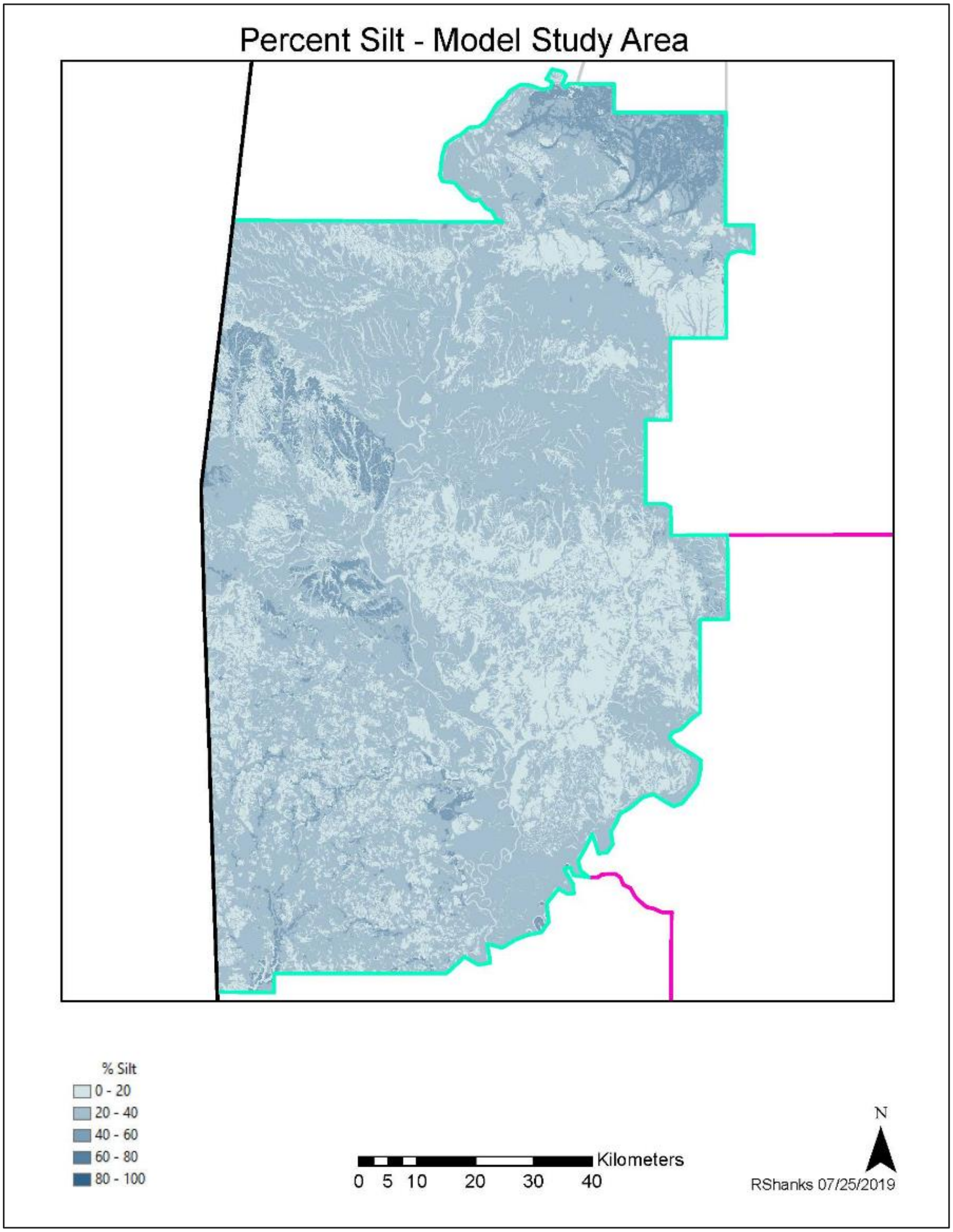

Percent Clay - Test Area 2
% Clay
0 - 20
20 - 40
40 - 60
60 - 80
80 - 100
0 5 10 20 30 40
Kilometers
N
RShanks 07/25/2019


Percent Sand - Model Study Area
% Sand
0 - 20
20 - 40
40 - 60
60 - 80
80 - 100
Kilometers
0 5 10 20 30 40
N
RShanks 07/25/2019

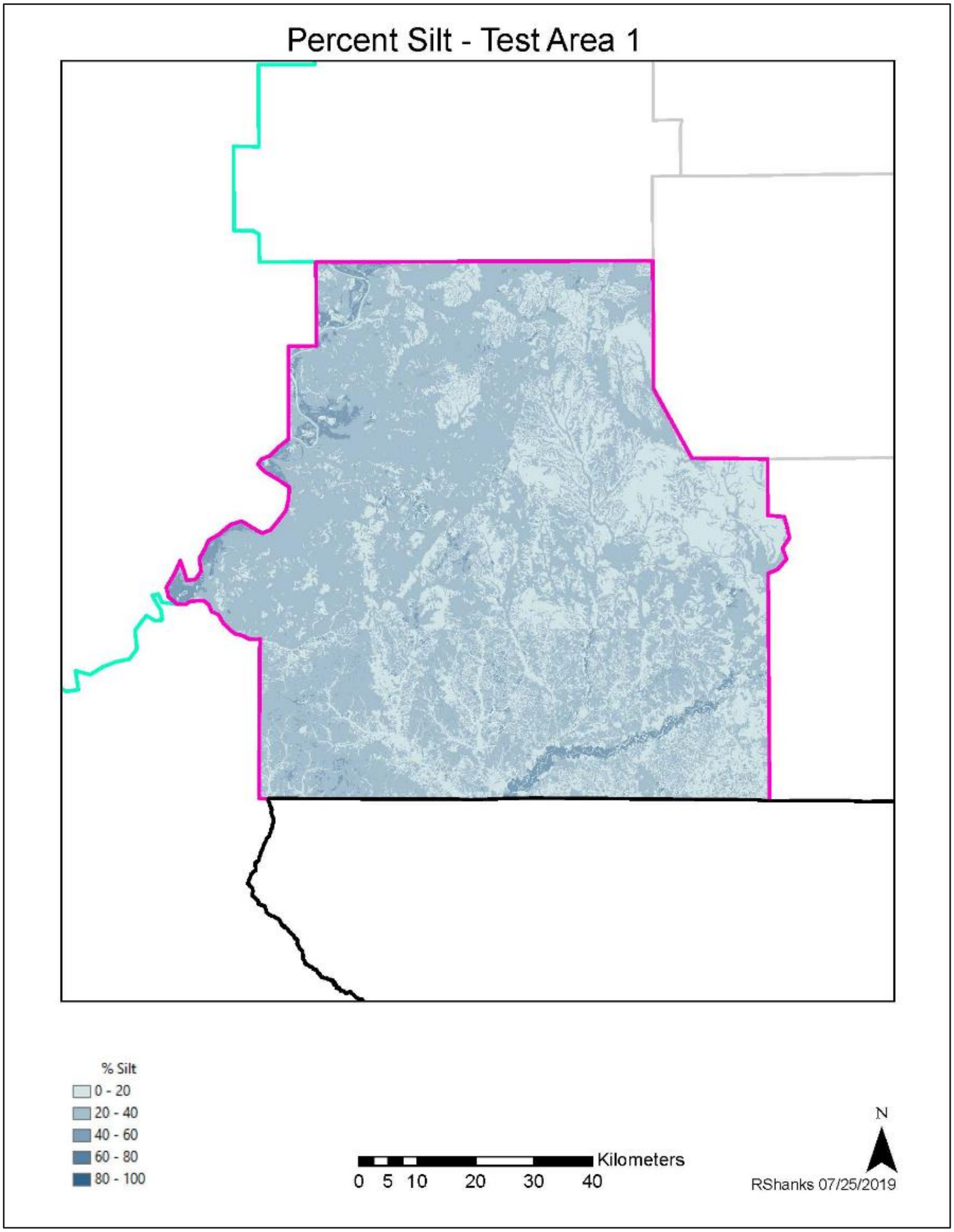

Percent Sand - Test Area 1
% Sand
0 - 20
20 - 40
40 - 60
60 - 80
80 - 100
0 5 10 20 30 40
Kilometers
N
RShanks 07/25/2019

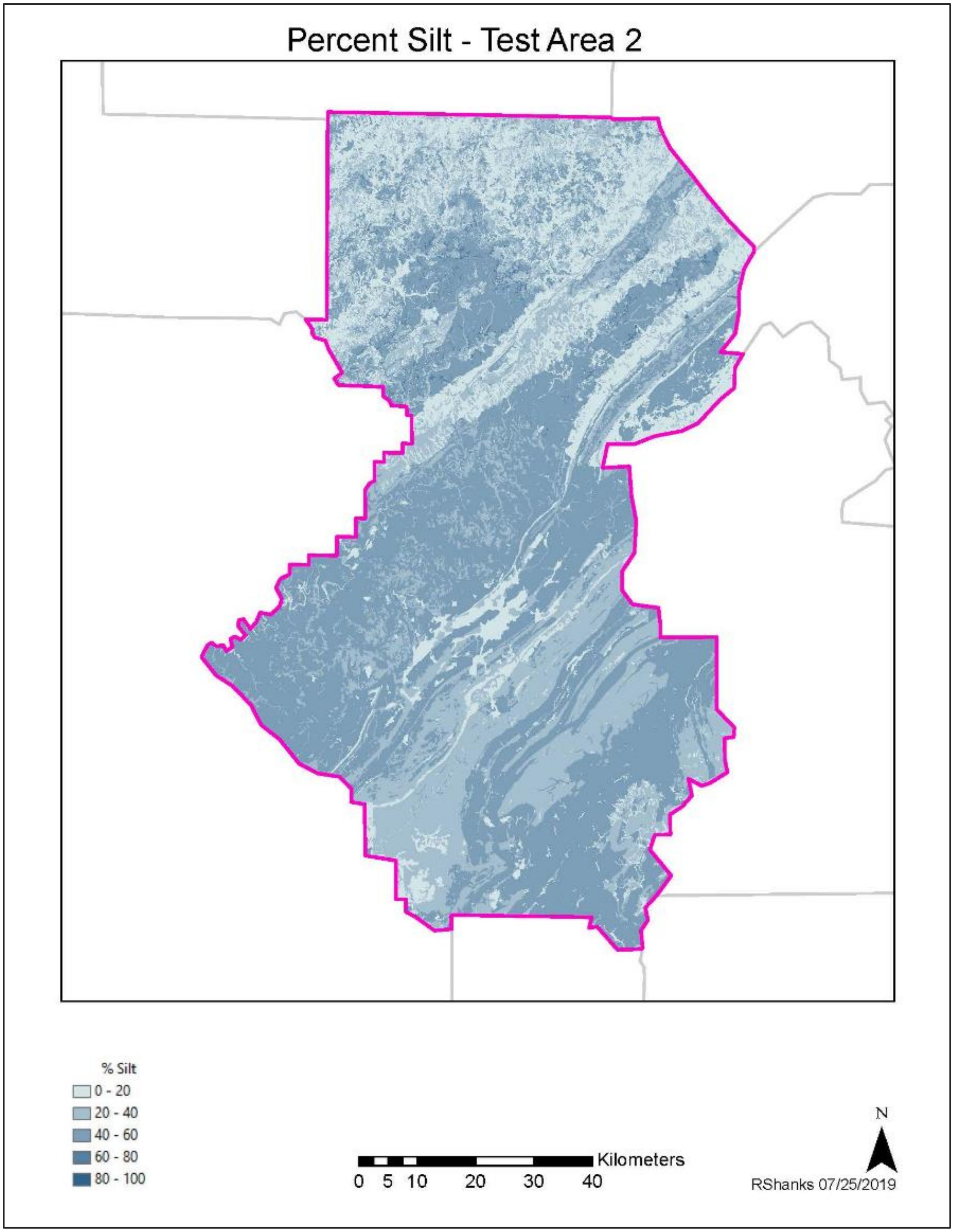


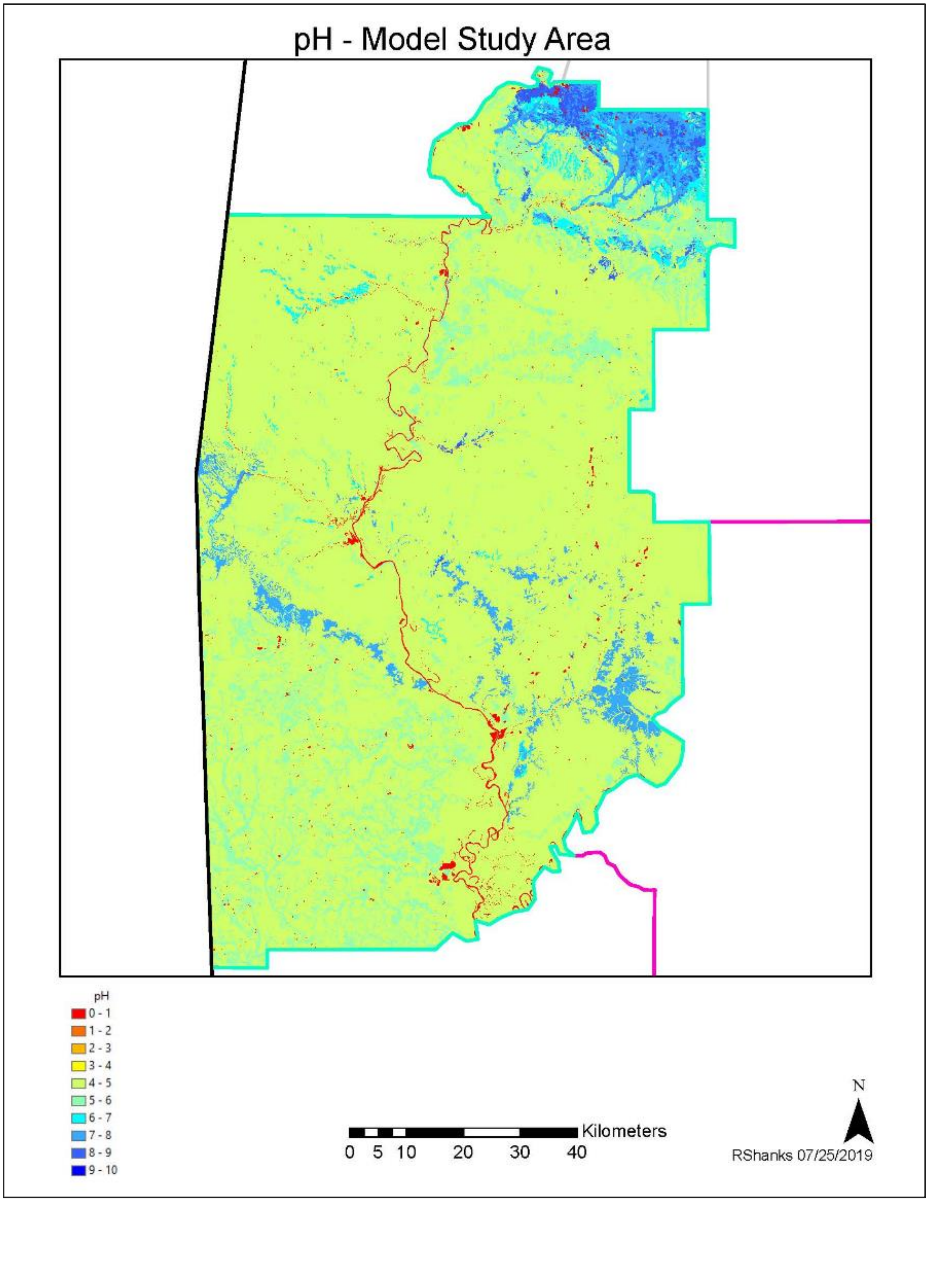
Percent Sand - Test Area 2
% Sand
0 - 20
20 - 40
40 - 60
60 - 80
80 - 100
0 5 10 20 30 40
Kilometers
N
RShanks 07/25/2019

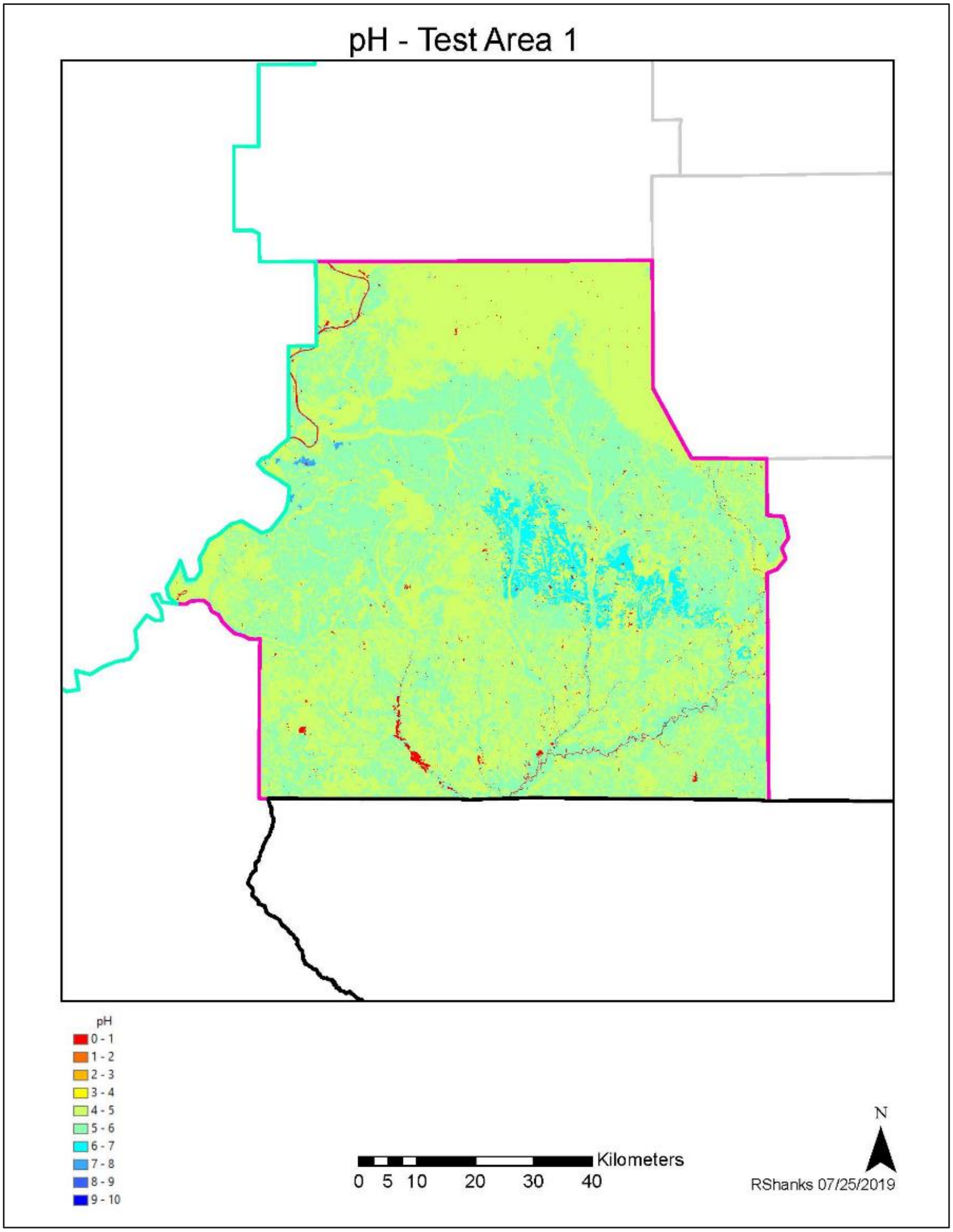


Percent Silt - Model Study Area
% Silt
0 - 20
20 - 40
40 - 60
60 - 80
80 - 100
Kilometers
0 5 10 20 30 40
N
RShanks 07/25/2019

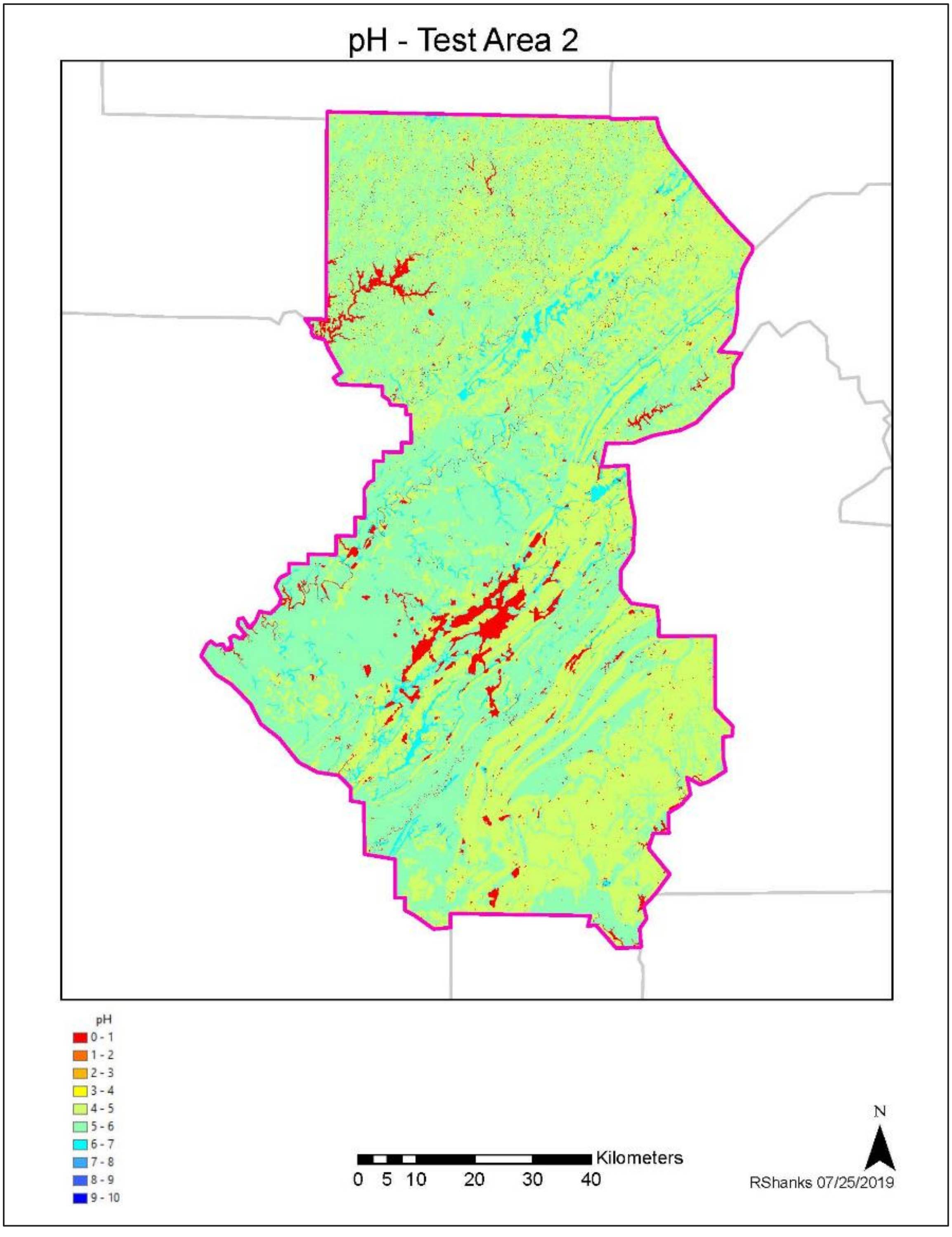

Percent Silt - Test Area 1
% Silt
0 - 20
20 - 40
40 - 60
60 - 80
80 - 100
Kilometers
0 5 10 20 30 40
N
RShanks 07/25/2019


Percent Silt - Test Area 2
% Silt
0 - 20
20 - 40
40 - 60
60 - 80
80 - 100
0 5 10 20 30 40 Kilometers
N
RShanks 07/25/2019

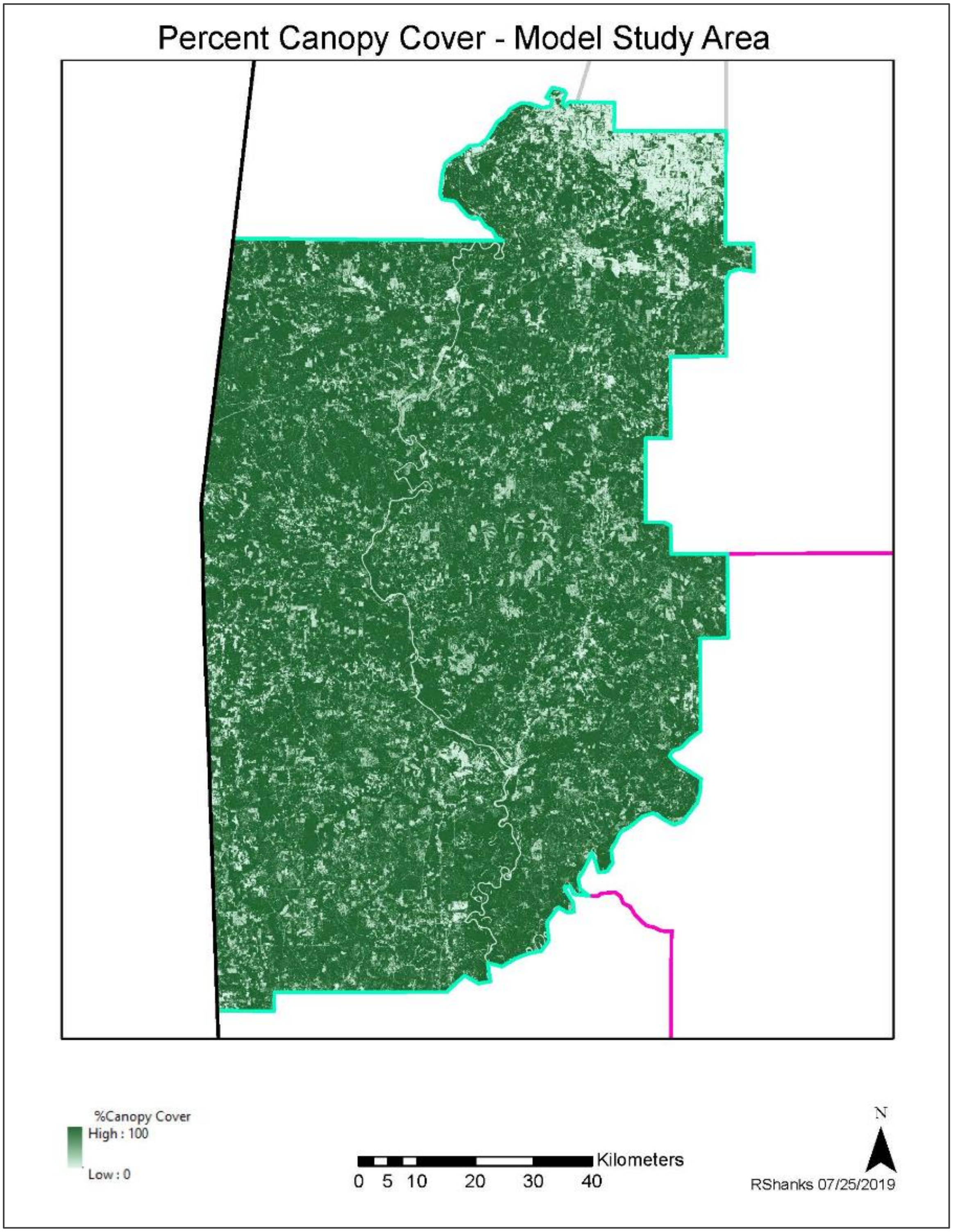

pH - Model Study Area
pH
0 - 1
1 - 2
2 - 3
3 - 4
4 - 5
5 - 6
6 - 7
7 - 8
8 - 9
9 - 10
Kilometers
0 5 10 20 30 40
N
RShanks 07/25/2019

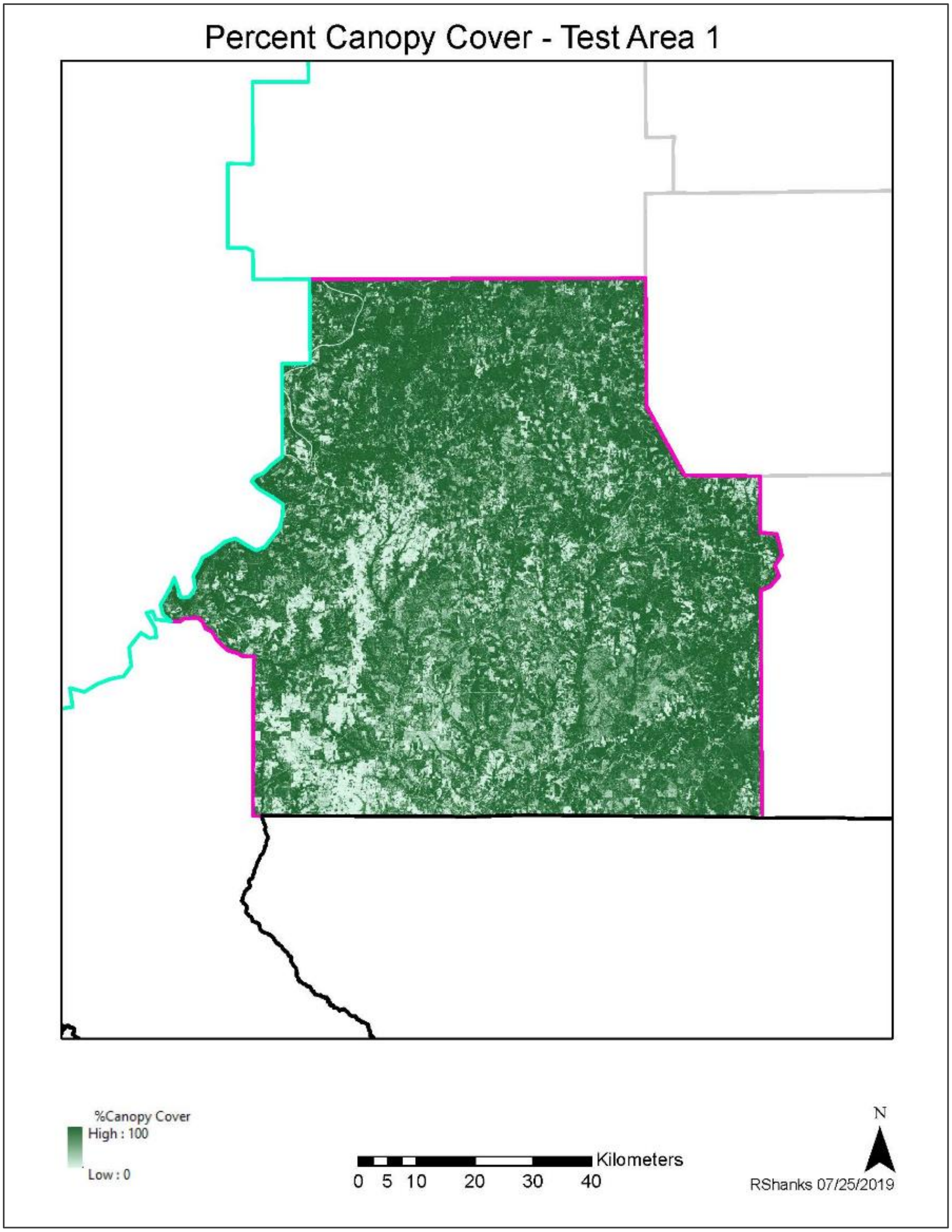


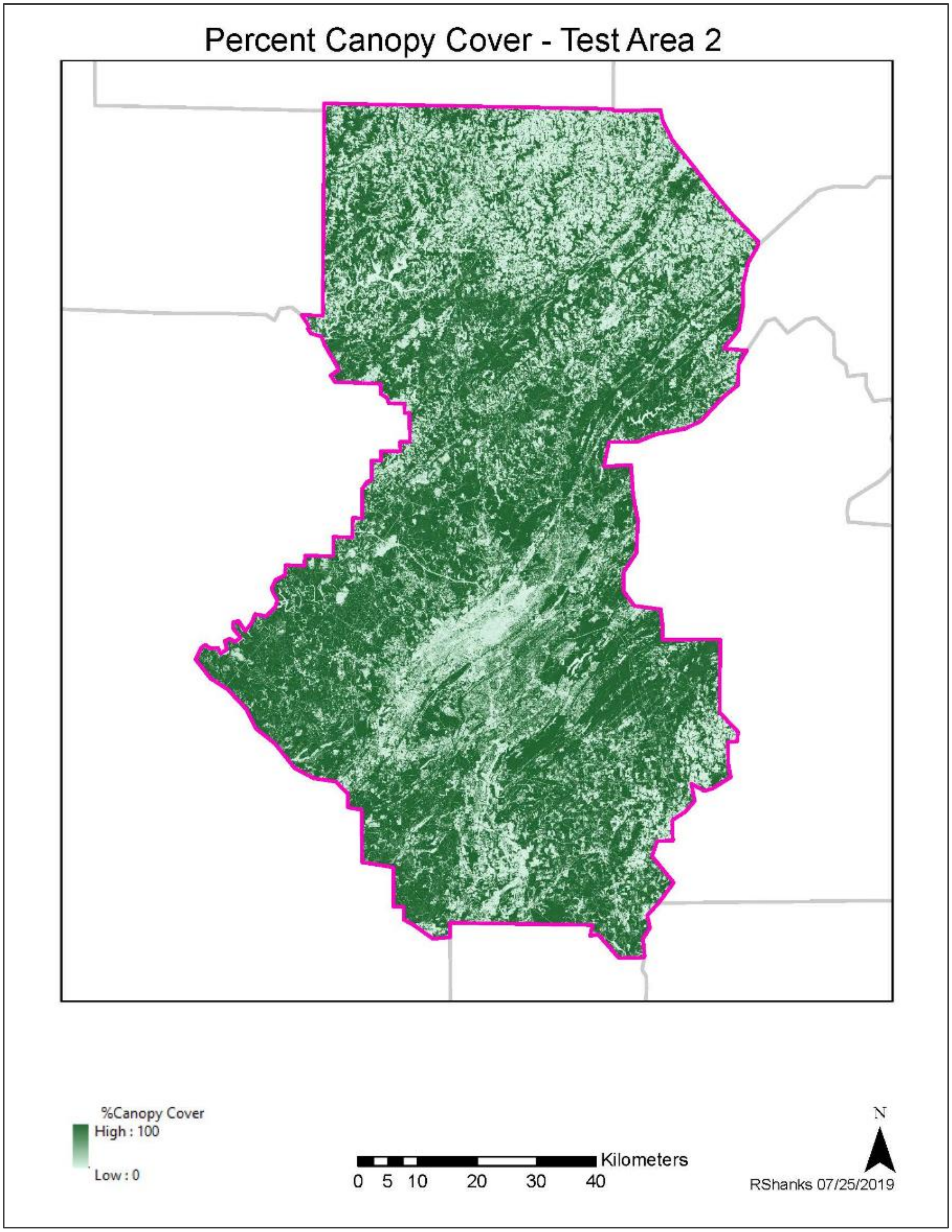
pH - Test Area 1
pH
0 - 1
1 - 2
2 - 3
3 - 4
4 - 5
5 - 6
6 - 7
7 - 8
8 - 9
9 - 10
0 5 10 20 30 40
Kilometers
N
RShanks 07/25/2019


pH - Test Area 2
pH
0 - 1
1 - 2
2 - 3
3 - 4
4 - 5
5 - 6
6 - 7
7 - 8
8 - 9
9 - 10
Kilometers
0 5 10 20 30 40
N
RShanks 07/25/2019

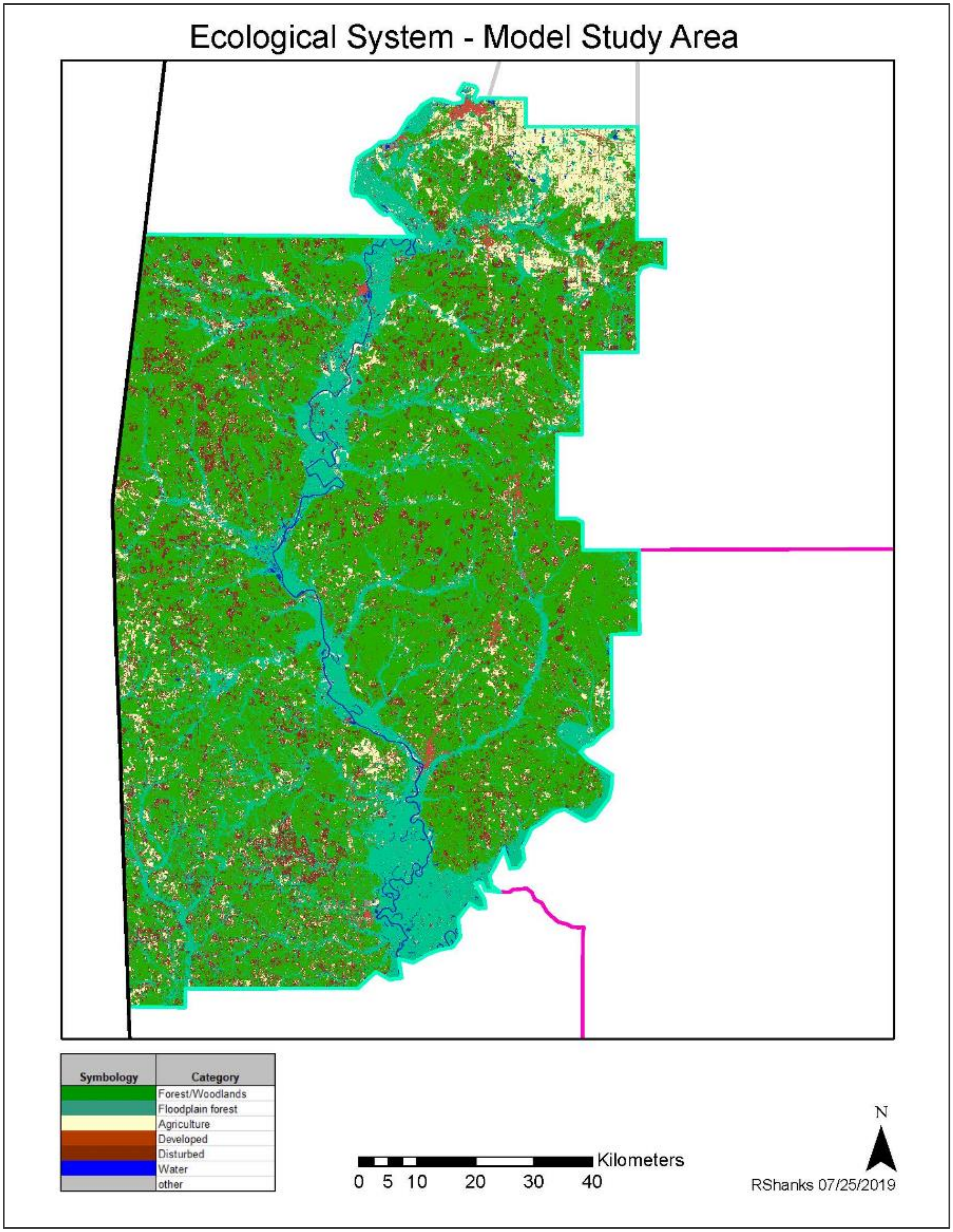
Appendix B: Other Environmental Covariate Maps

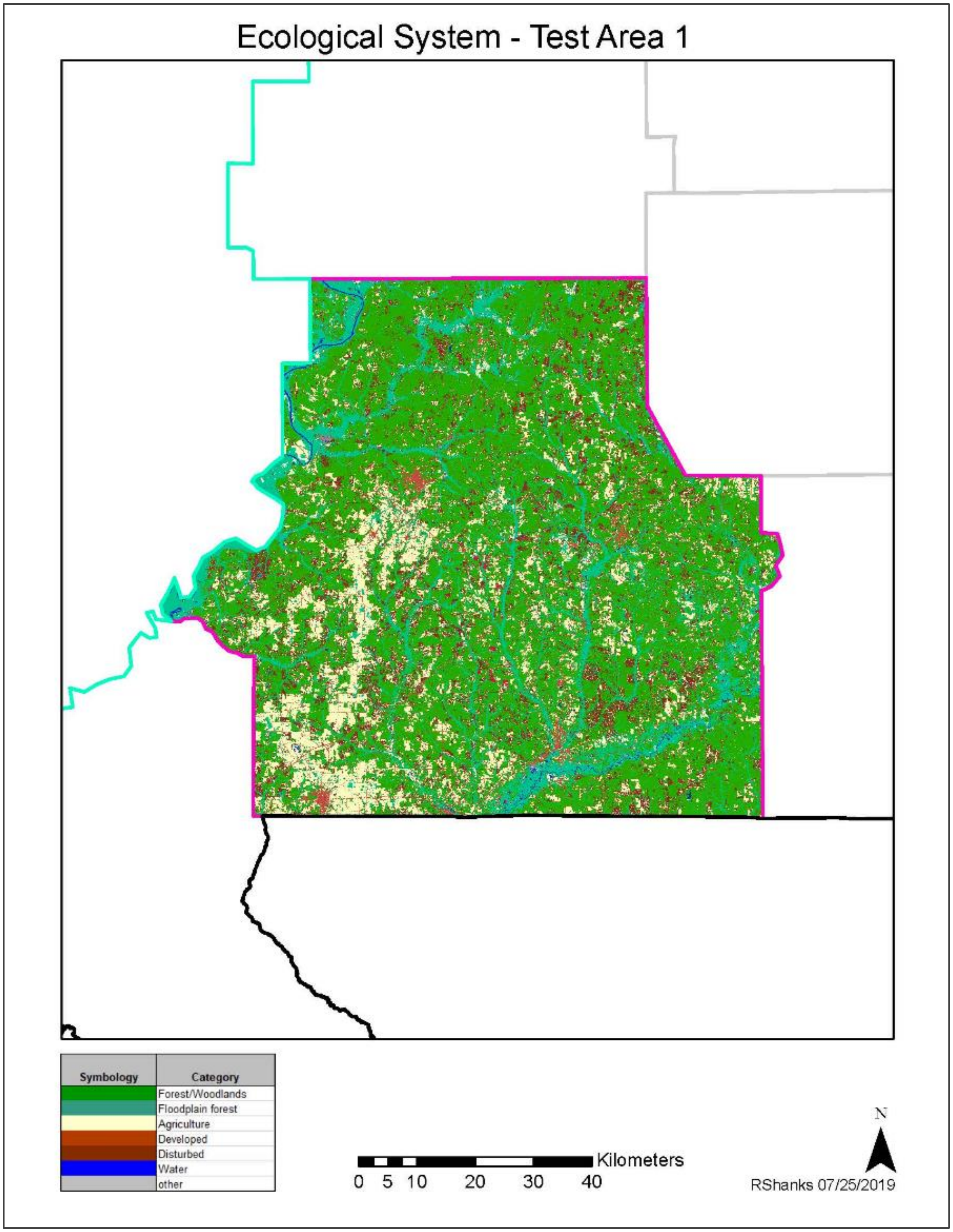


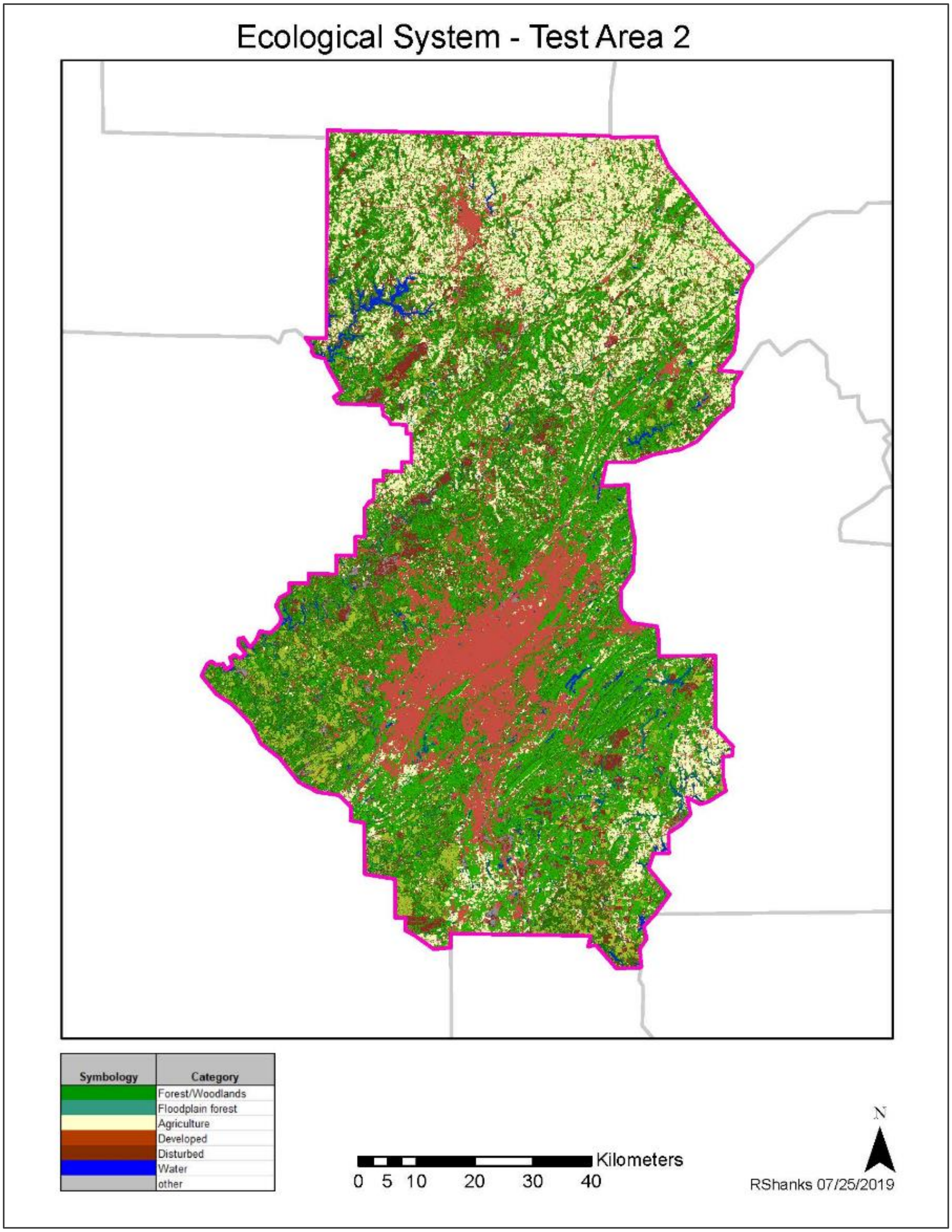

Percent Canopy Cover - Test Area 1
%Canopy Cover
High : 100
Low : 0
Kilometers
0 5 10 20 30 40
N
RShanks 07/25/2019

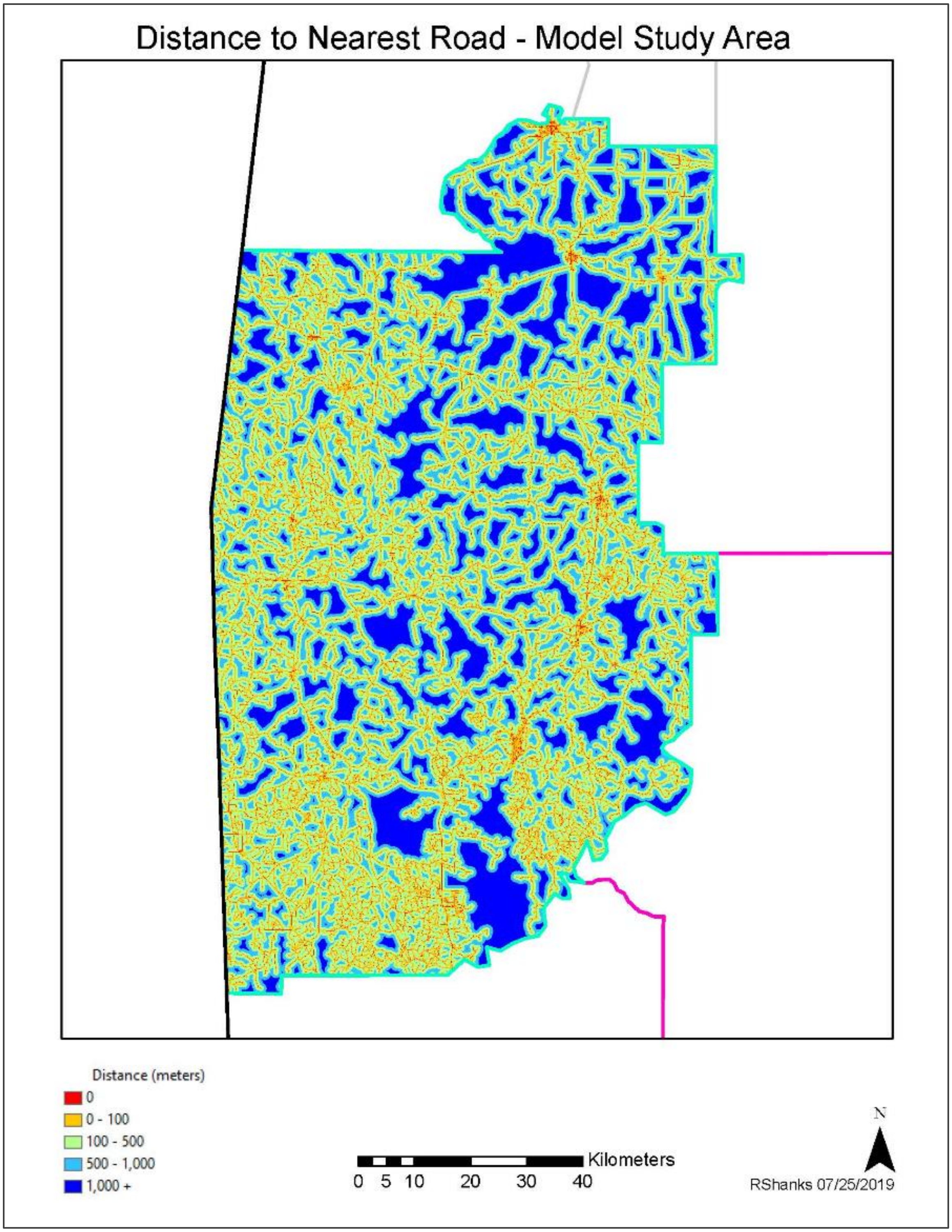


Percent Canopy Cover - Test Area 2
%Canopy Cover
High : 100
Low : 0
Kilometers
0 5 10 20 30 40
N
RShanks 07/25/2019

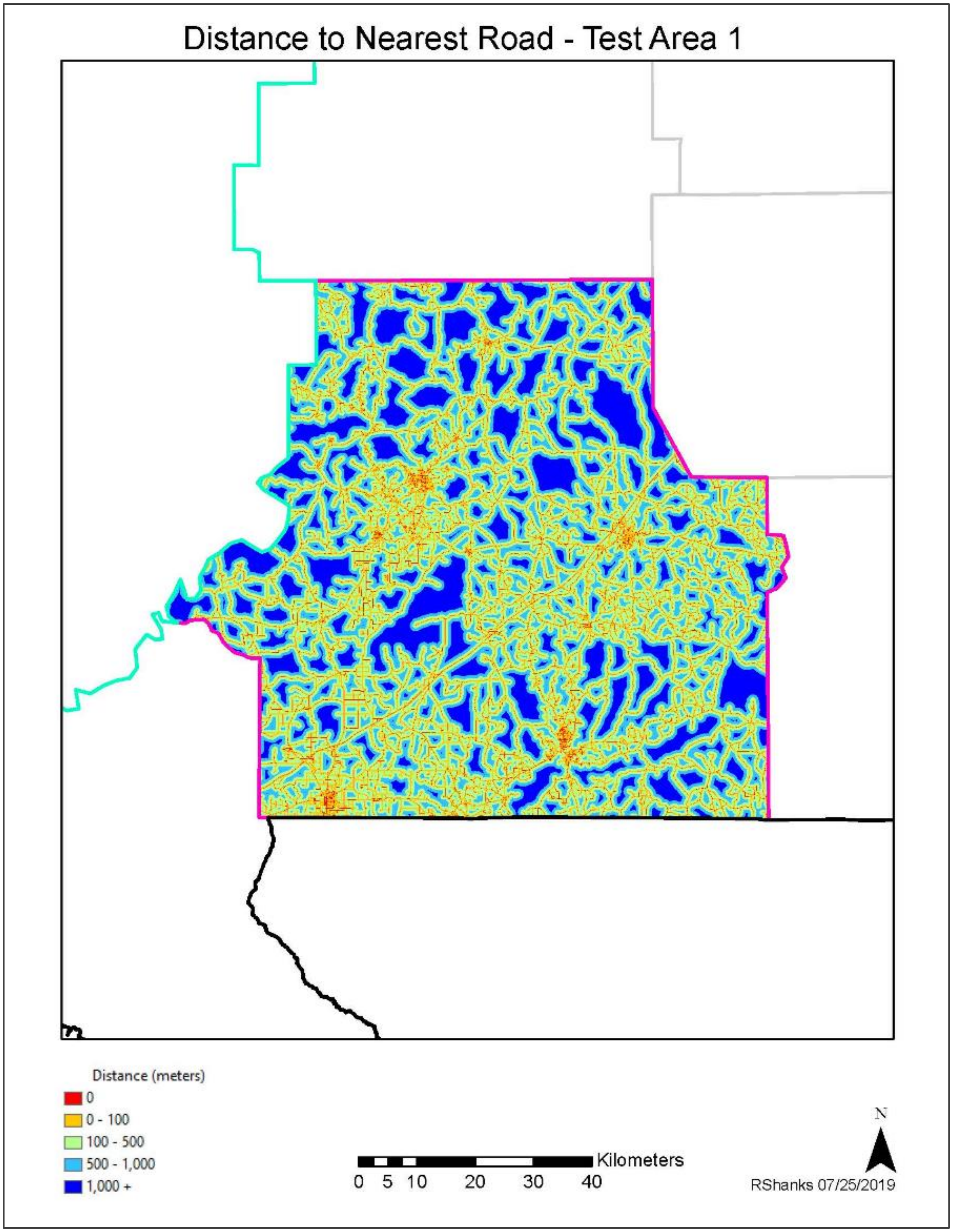

Ecological System - Model Study Area
Symbology
Category
Forest/Woodlands
Floodplain forest
Agriculture
Developed
Disturbed
Water
other
Kilometers
0 5 10 20 30 40
N
RShanks 07/25/2019

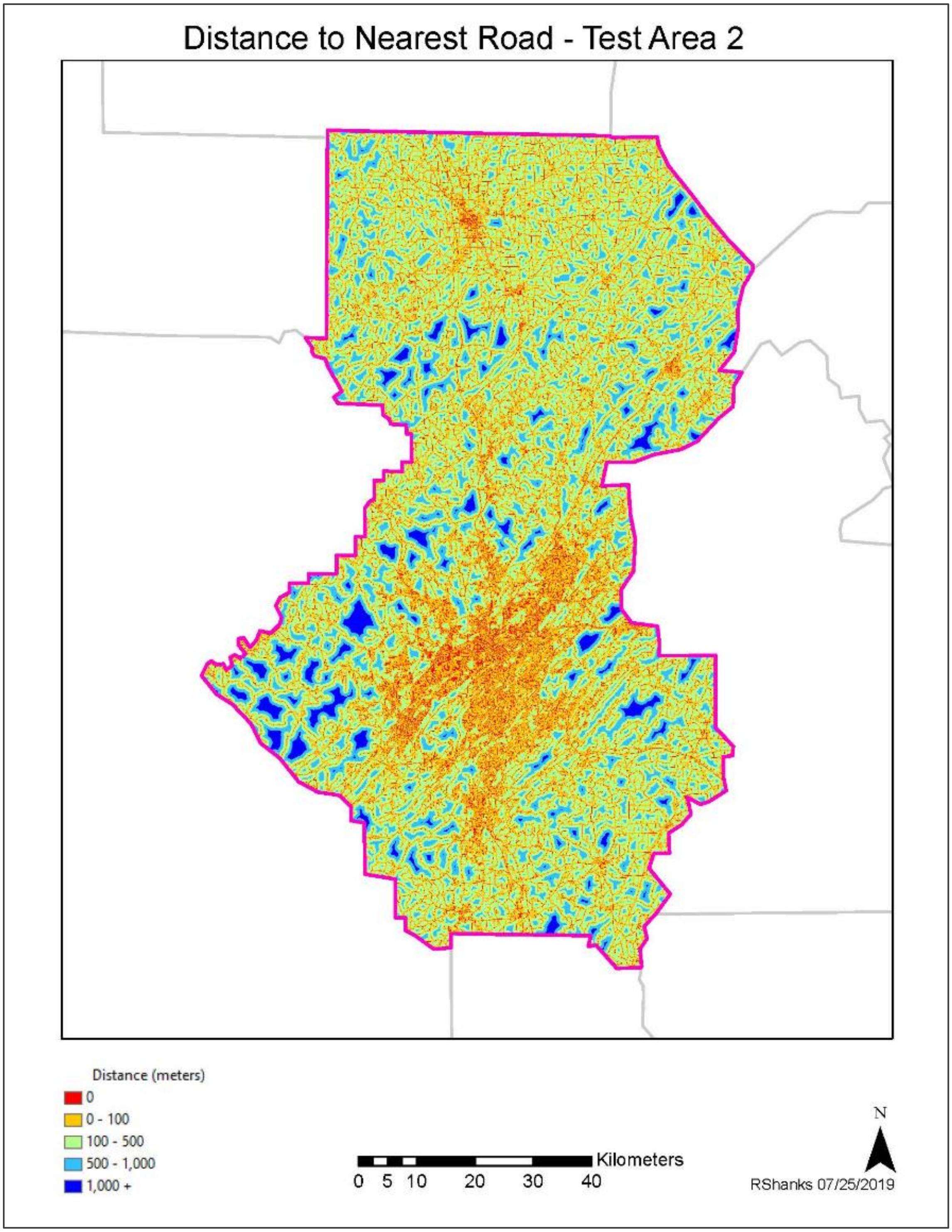

Ecological System - Test Area 1
Symbology
Category
Forest/Woodlands
Floodplain forest
Agriculture
Developed
Disturbed
Water
other
Kilometers
0 5 10 20 30 40
N
RShanks 07/25/2019


Ecological System - Test Area 2
Symbology
Category
Forest/Woodlands
Floodplain forest
Agriculture
Developed
Disturbed
Water
other
Kilometers
0 5 10 20 30 40
N
RShanks 07/25/2019



Distance to Nearest Road - Model Study Area
Distance (meters)
0
0 - 100
100 - 500
500 - 1,000
1,000 +
Kilometers
0 5 10 20 30 40
N
RShanks 07/25/2019


Distance to Nearest Road - Test Area 1
Distance (meters)
0
0 - 100
100 - 500
500 - 1,000
1,000 +
0 5 10 20 30 40
Kilometers
N
RShanks 07/25/2019


Distance to Nearest Road - Test Area 2
Distance (meters)
0
0 - 100
100 - 500
500 - 1,000
1,000 +
Kilometers
0 5 10 20 30 40
N
RShanks 07/25/2019

Appendix C: Data Layer Conversion Steps

This table represents the step for converting the presence point data to .csv for use in Maxent

Step	Description
1	Open the cogongrass points shapefile in ArcGIS
2	Add new columns for LatYDD and LongXDD to the shapefile table
3	Calculate the Latitude and Longitude geometry as decimal degrees
4	Use the Table to Excel tool in ArcGIS to dump the data into Excel format
5	Open the file in Microsoft Excel
6	Convert the .xls file to .csv
7	Open the .csv file
8	Remove all columns except Species, LatYDD, and LongXDD
9	Save the file

This table of conversion steps includes data sources and layers that were ultimately not used in the final model; however, it may be useful for the reader to review how data from these sources were prepared before their usefulness was determined to be insignificant to the study.

<table>
<tr><th>Environmental Layer</th><th>Description and Conversion Steps</th></tr>
<tr><td>PRISM climate data</td><td>

Download climate data from the PRISM Climate Group website (http://www.prism.oregonstate.edu/)
Each dataset is a raster dataset at 800M resolution (roughly ½ mile grid cells). The values in the dataset are presented in millimeters and the rasters are classified in 5-inch increments. These datasets are in Nad83.
Data conversion steps:

In ArcGIS

Open the AVG Precipitation dataset

Clip AVG Precipitation to the study area geometry (the state of Alabama)
Save as “AVGPrecipClip”

Open the AVG Min Temp dataset

Clip AVG Min Temp to the study area geometry (the state of Alabama)
Save as “AVGMinTempClip”

Open the AVG Max Temp dataset

Clip AVG Max Temp to the study area geometry (the state of Alabama)
Save as “AVGMaxTempClip”

</td></tr>
<tr><td>Digital Elevation Model</td><td>

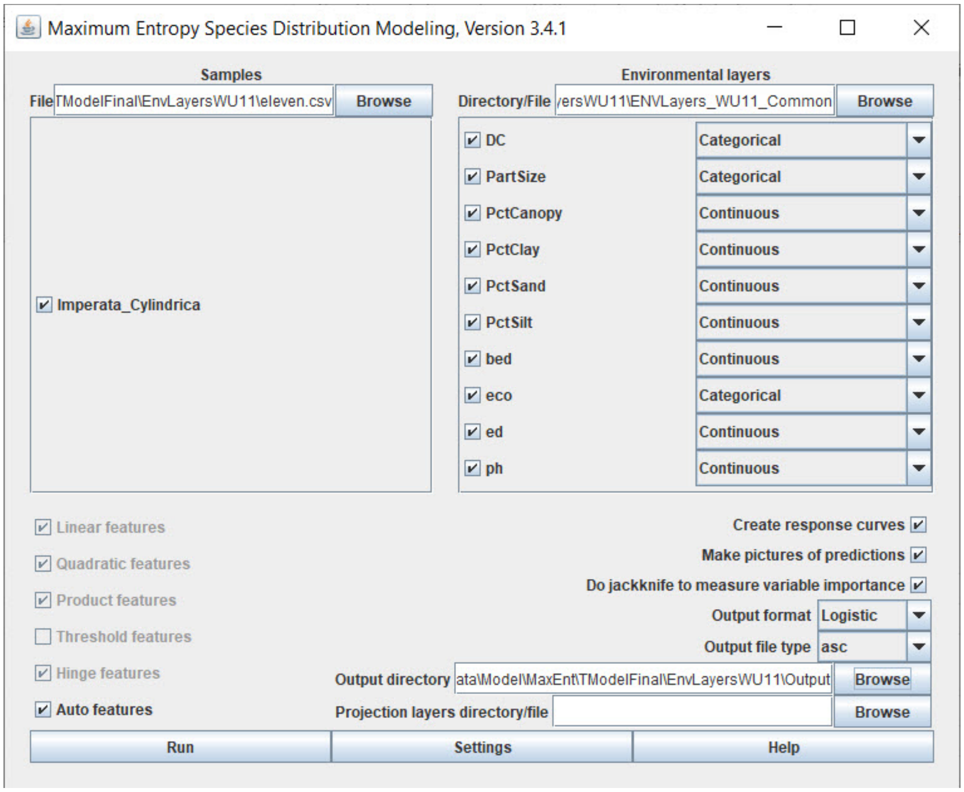
Download digital elevation models for the state of Alabama
In ArcGIS
Use the Merge to New Raster tool to merge the DEMs together into one raster

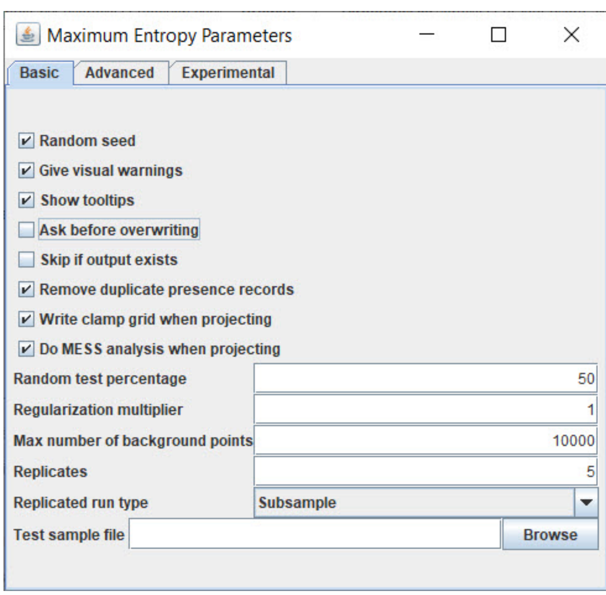
Use Clip Raster tool to clip the new raster to the study area

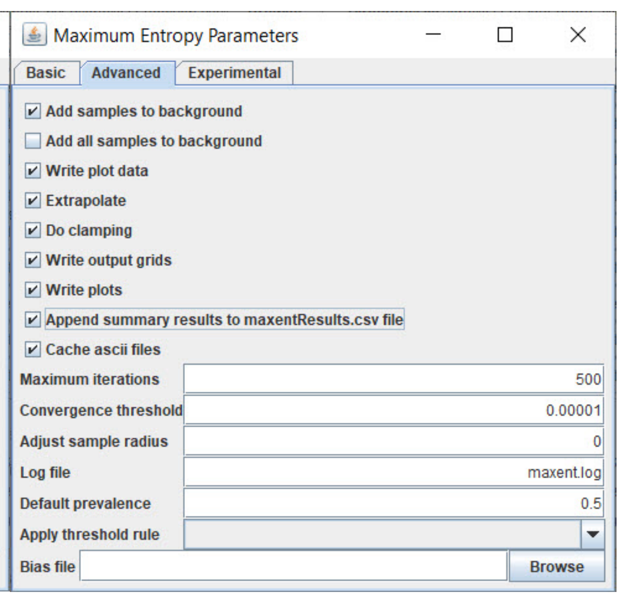
</td></tr>
</table>

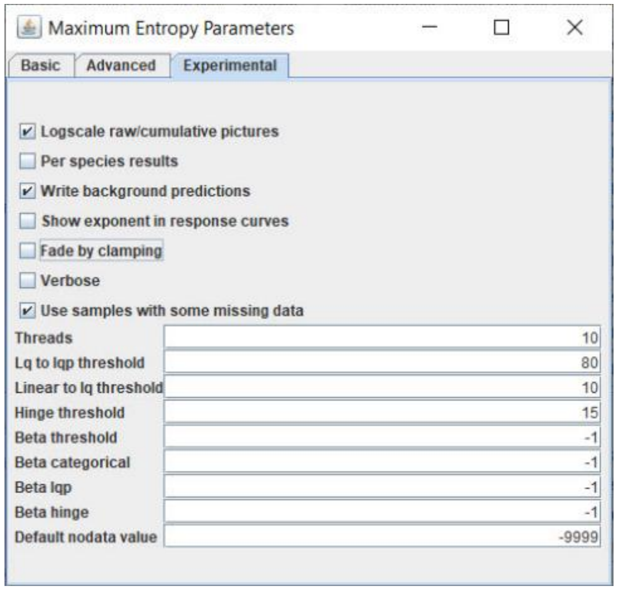
Appendix D: Maxent Model Settings Screen Captures

Model Study Area Maxent Model Settings:

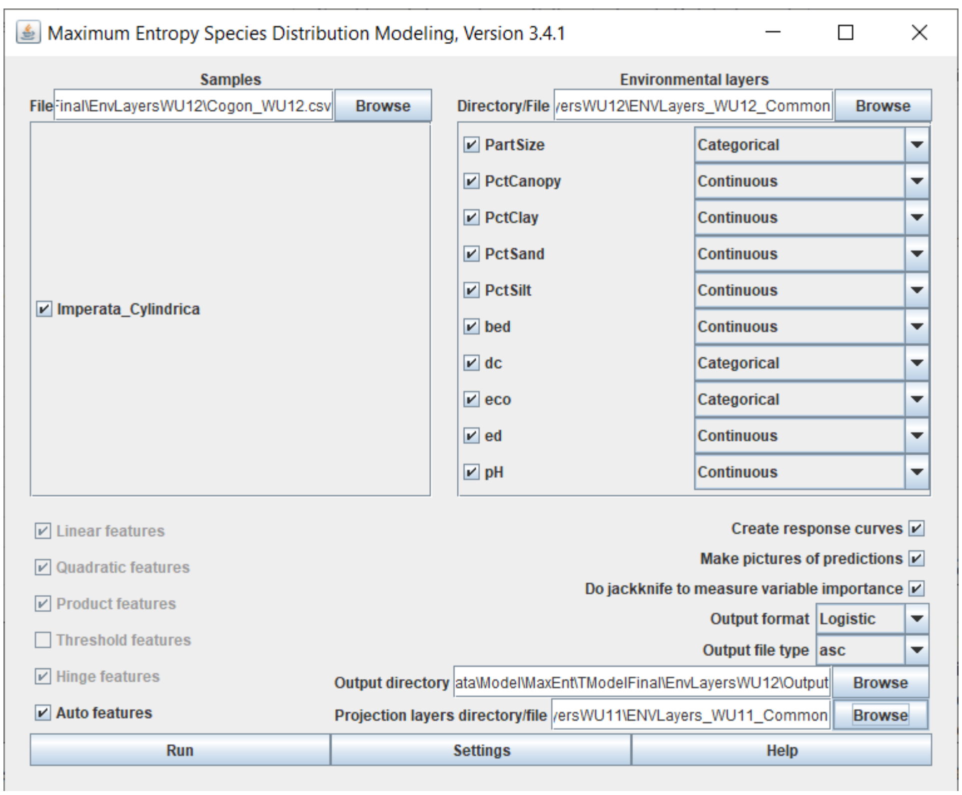




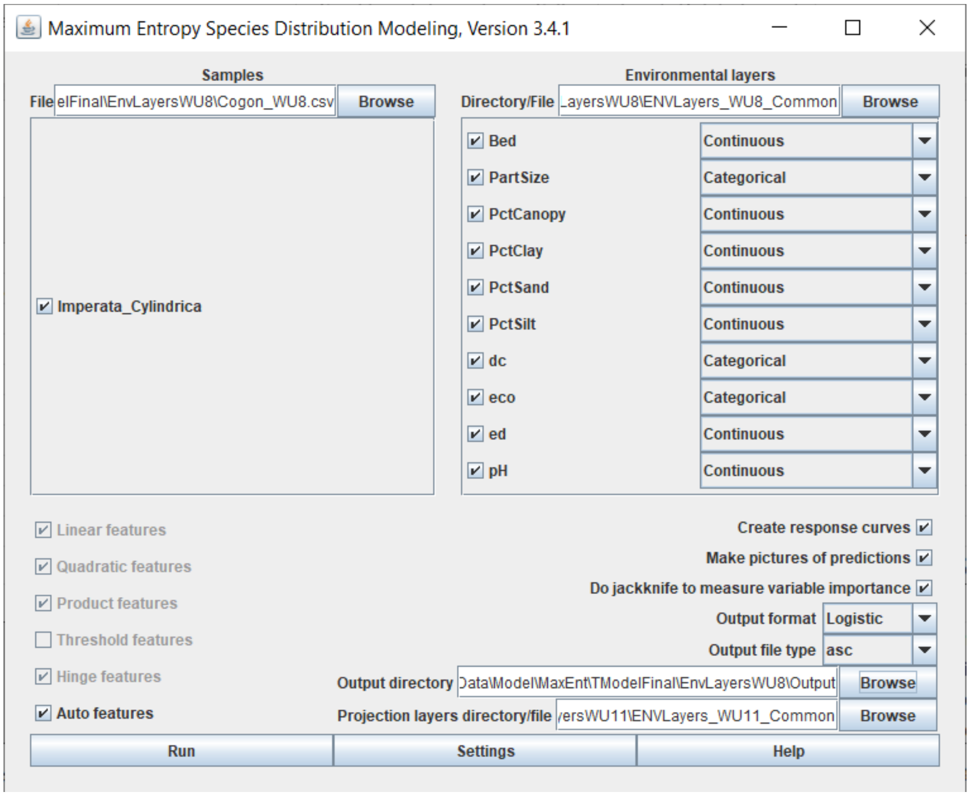




Test Study Area 1 (Work Unit 12) Maxent Model Settings:

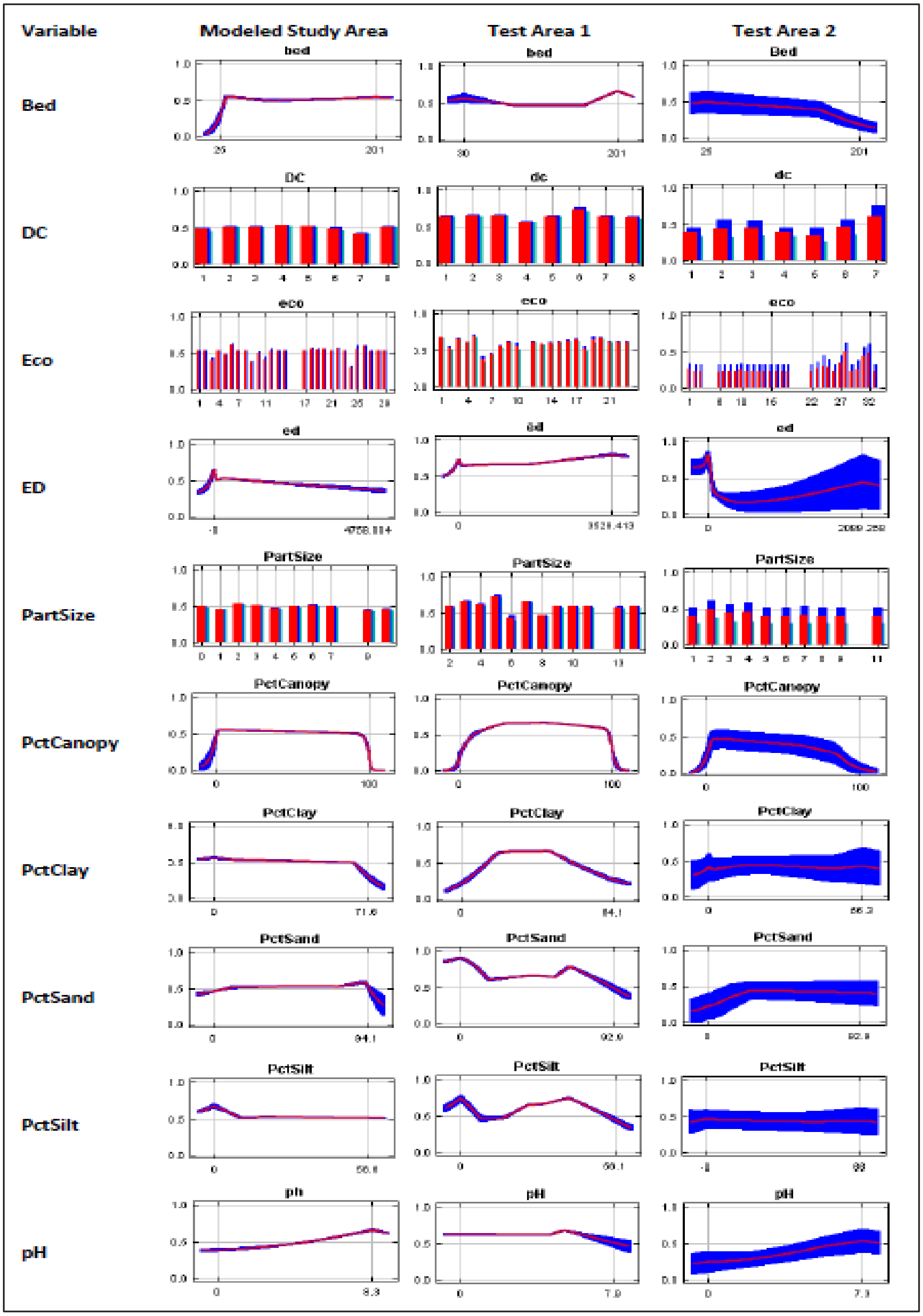


Test Study Area 2 (Work Unit 8) Maxent Model Settings:



Appendix E: Response Curves

Response Curves for *Imperata cylindrica* to each environmental variable included in the models for each of the three study areas.



Appendix F: Ecological Systems with Category Groupings

Ecological Systems for Model Study Area:

ID	Ecological System	% of Total	Category
1	Cultivated Cropland	1.21%	Agriculture
2	Developed, High Intensity	0.03%	Developed
3	Developed, Low Intensity	0.26%	Developed
4	Developed, Medium Intensity	0.09%	Developed
5	Developed, Open Space	2.80%	Developed
6	Disturbed/Successional - Shrub Regeneration	1.83%	Disturbed
7	East Gulf Coastal Plain Black Belt Calcareous Prairie and Woodland - Herbaceous Modifier	0.02%	Forest/Woodlands
8	East Gulf Coastal Plain Black Belt Calcareous Prairie and Woodland - Woodland Modifier	0.08%	Forest/Woodlands
9	East Gulf Coastal Plain Dry Chalk Bluff	0.00%	other
10	East Gulf Coastal Plain Interior Upland Longleaf Pine Woodland - Loblolly Modifier	34.12%	Forest/Woodlands
11	East Gulf Coastal Plain Interior Upland Longleaf Pine Woodland - Offsite Hardwood Modifier	8.37%	Forest/Woodlands
12	East Gulf Coastal Plain Interior Upland Longleaf Pine Woodland - Open Understory Modifier	0.71%	Forest/Woodlands
13	East Gulf Coastal Plain Large River Floodplain Forest - Forest Modifier	8.83%	Floodplain forest
14	East Gulf Coastal Plain Large River Floodplain Forest - Herbaceous Modifier	0.16%	Floodplain forest
15	East Gulf Coastal Plain Limestone Forest	0.09%	Forest/Woodlands
16	East Gulf Coastal Plain Northern Mesic Hardwood Forest	0.10%	Floodplain forest
17	East Gulf Coastal Plain Small Stream and River Floodplain Forest	8.11%	Floodplain forest
18	East Gulf Coastal Plain Southern Loblolly-Hardwood Flatwoods	0.94%	Forest/Woodlands
19	East Gulf Coastal Plain Southern Mesic Slope Forest	5.72%	Floodplain forest
20	Evergreen Plantation or Managed Pine	5.66%	Forest/Woodlands
21	Harvested Forest-Shrub Regeneration	7.54%	Disturbed
22	Harvested Forest - Grass/Forb Regeneration	1.61%	Disturbed
23	Open Water (Aquaculture)	0.03%	water
24	Open Water (Fresh)	1.15%	water
25	Pasture/Hay	6.86%	Agriculture
26	Quarries, Mines, Gravel Pits and Oil Wells	0.01%	Developed
27	Southern Coastal Plain Blackwater River Floodplain Forest	3.61%	Floodplain forest
28	Unconsolidated Shore	0.01%	other
29	Undifferentiated Barren Land	0.06%	other

Ecological Systems for Study Area 1:

ID	Ecological System Description	% of Total	Category
1	Cultivated Cropland	7.67%	Agriculture
2	Developed, High Intensity	0.03%	Developed
3	Developed, Low Intensity	0.53%	Developed
4	Developed, Medium Intensity	0.10%	Developed
5	Developed, Open Space	2.36%	Developed
6	Disturbed/Successional - Shrub Regeneration	1.99%	Disturbed
7	East Gulf Coastal Plain Interior Upland Longleaf Pine Woodland - Loblolly Modifier	29.35%	Forest/Woodlands
8	East Gulf Coastal Plain Interior Upland Longleaf Pine Woodland - Offsite Hardwood Modifier	8.78%	Forest/Woodlands
9	East Gulf Coastal Plain Interior Upland Longleaf Pine Woodland - Open Understory Modifier	2.68%	Forest/Woodlands
10	East Gulf Coastal Plain Large River Floodplain Forest - Forest Modifier	4.33%	Floodplain forest
11	East Gulf Coastal Plain Large River Floodplain Forest - Herbaceous Modifier	0.16%	Floodplain forest
12	East Gulf Coastal Plain Small Stream and River Floodplain Forest	5.12%	Floodplain forest
13	East Gulf Coastal Plain Southern Mesic Slope Forest	6.94%	Floodplain forest
14	Evergreen Plantation or Managed Pine	6.07%	Forest/Woodlands
15	Harvested Forest-Shrub Regeneration	7.25%	Disturbed
16	Harvested Forest - Grass/Forb Regeneration	3.15%	Disturbed
17	Open Water (Fresh)	0.68%	Water
18	Pasture/Hay	6.97%	Agriculture
19	Quarries, Mines, Gravel Pits and Oil Wells	0.07%	Developed
20	Southern Coastal Plain Blackwater River Floodplain Forest	5.56%	Floodplain forest
21	Southern Coastal Plain Nonriverine Cypress Dome	0.13%	Floodplain forest
22	Unconsolidated Shore	0.01%	other
23	Undifferentiated Barren Land	0.07%	other

Ecological Systems for Study Area 2:

ID	Ecological System Description	% of Total	Category
1	Allegheny-Cumberland Dry Oak Forest and Woodland - Hardwood	7.20%	Forest/Woodlands
2	Allegheny-Cumberland Dry Oak Forest and Woodland - Pine Modifier	0.92%	Forest/Woodlands
3	Cultivated Cropland	3.38%	Agriculture
4	Cumberland Riverscour	0.16%	water
5	Developed, High Intensity	0.53%	Developed
6	Developed, Low Intensity	4.48%	Developed
7	Developed, Medium Intensity	1.32%	Developed
8	Developed, Open Space	8.73%	Developed
9	Disturbed/Successional - Grass/Forb Regeneration	1.87%	Disturbed
10	Disturbed/Successional - Shrub Regeneration	3.05%	Disturbed
11	Evergreen Plantation or Managed Pine	5.36%	Forest/Woodlands
12	Harvested Forest-Shrub Regeneration	1.72%	Disturbed
13	Harvested Forest - Grass/Forb Regeneration	1.56%	Disturbed
14	Northeastern Interior Dry Oak Forest - Mixed Modifier	0.00%	Forest/Woodlands
15	Open Water (Fresh)	1.65%	water
16	Pasture/Hay	17.46%	Agriculture
17	South-Central Interior Large Floodplain - Forest Modifier	0.09%	floodplain forest
18	South-Central Interior Mesophytic Forest	6.88%	Forest/Woodlands
19	South-Central Interior Small Stream and Riparian	0.99%	water
20	Southeastern Interior Longleaf Pine Woodland	0.29%	Forest/Woodlands
21	Southern Appalachian Low Mountain Pine Forest	8.70%	Forest/Woodlands
22	Southern Interior Acid Cliff	0.00%	other
23	Southern Interior Calcareous Cliff	0.00%	other
24	Southern Interior Low Plateau Dry-Mesic Oak Forest	0.00%	Forest/Woodlands
25	Southern Piedmont Cliff	0.00%	Other
26	Southern Piedmont Dry Oak-(Pine) Forest - Hardwood Modifier	0.64%	Forest/Woodlands
27	Southern Piedmont Dry Oak-(Pine) Forest - Loblolly Pine Modifier	0.08%	Forest/Woodlands
28	Southern Piedmont Dry Oak-(Pine) Forest - Mixed Modifier	0.09%	Forest/Woodlands
29	Southern Piedmont Mesic Forest	0.11%	Forest/Woodlands
30	Southern Piedmont Small Floodplain and Riparian Forest	0.04%	floodplain forest
31	Southern Ridge and Valley Dry Calcareous Forest	20.74%	Forest/Woodlands
32	Southern Ridge and Valley Dry Calcareous Forest - Pine modifier	1.32%	Forest/Woodlands
33	Undifferentiated Barren Land	0.64%	Other

Ecological Systems for the Model Study Area, Test Area 1, and Test Area 2 with consolidated groupings.

OBJECTID	ECOLSYS_LU	MSA	TA 1	TA 2	category
1	Allegheny-Cumberland Dry Oak Forest and Woodland - Hardwood	0.00%	0.00%	7.20%	Forest/Woodlands
2	Allegheny-Cumberland Dry Oak Forest and Woodland - Pine Modifier	0.00%	0.00%	0.92%	Forest/Woodlands
3	Cultivated Cropland	1.21%	7.67%	3.38%	Agriculture
4	Cumberland Riverscour	0.00%	0.00%	0.16%	water
5	Developed, High Intensity	0.03%	0.03%	0.53%	Developed
6	Developed, Low Intensity	0.26%	0.53%	4.48%	Developed
7	Developed, Medium Intensity	0.09%	0.10%	1.32%	Developed
8	Developed, Open Space	2.80%	2.36%	8.73%	Developed
9	Disturbed/Successional - Grass/Forb Regeneration	0.00%	0.00%	1.87%	Disturbed
10	Disturbed/Successional - Shrub Regeneration	1.83%	1.99%	3.05%	Disturbed
11	East Gulf Coastal Plain Black Belt Calcareous Prairie and Woodland - Herbaceous Modifier	0.02%	0.00%	0.00%	Forest/Woodlands
12	East Gulf Coastal Plain Black Belt Calcareous Prairie and Woodland - Woodland Modifier	0.08%	0.00%	0.00%	Forest/Woodlands
13	East Gulf Coastal Plain Dry Chalk Bluff	0.00%	0.00%	0.00%	other
14	East Gulf Coastal Plain Interior Upland Longleaf Pine Woodland - Loblolly Modifier	34.12%	29.35%	0.00%	Forest/Woodlands
15	East Gulf Coastal Plain Interior Upland Longleaf Pine Woodland - Offsite Hardwood Modifier	8.37%	8.78%	0.00%	Forest/Woodlands
16	East Gulf Coastal Plain Interior Upland Longleaf Pine Woodland - Open Understory Modifier	0.71%	2.68%	0.00%	Forest/Woodlands
17	East Gulf Coastal Plain Large River Floodplain Forest - Forest Modifier	8.83%	4.33%	0.00%	Floodplain forest
18	East Gulf Coastal Plain Large River Floodplain Forest - Herbaceous Modifier	0.16%	0.16%	0.00%	Floodplain forest
19	East Gulf Coastal Plain Limestone Forest	0.09%	0.00%	0.00%	Forest/Woodlands
20	East Gulf Coastal Plain Northern Mesic Hardwood Forest	0.10%	0.00%	0.00%	Floodplain forest
21	East Gulf Coastal Plain Small Stream and River Floodplain Forest	8.11%	5.12%	0.00%	Floodplain forest
22	East Gulf Coastal Plain Southern Loblolly-Hardwood Flatwoods	0.94%	0.00%	0.00%	Forest/Woodlands
23	East Gulf Coastal Plain Southern Mesic Slope Forest	5.72%	6.94%	0.00%	Floodplain forest
24	Evergreen Plantation or Managed Pine	5.66%	6.07%	5.36%	Forest/Woodlands
25	Harvested Forest-Shrub Regeneration	7.54%	7.25%	1.72%	Disturbed
26	Harvested Forest - Grass/Forb Regeneration	1.61%	3.15%	1.56%	Disturbed
27	Northeastern Interior Dry Oak Forest - Mixed Modifier	0.00%	0.00%	0.00%	Forest/Woodlands
28	Open Water (Aquaculture)	0.03%	0.00%	0.00%	water
29	Open Water (Fresh)	1.15%	0.68%	1.65%	water
30	Pasture/Hay	6.86%	6.97%	17.46%	Agriculture
31	Quarries, Mines, Gravel Pits and Oil Wells	0.01%	0.07%	0.00%	Developed
32	South-Central Interior Large Floodplain - Forest Modifier	0.00%	0.00%	0.09%	floodplain forest
33	South-Central Interior Mesophytic Forest	0.00%	0.00%	6.88%	Forest/Woodlands
34	South-Central Interior Small Stream and Riparian	0.00%	0.00%	0.99%	water
35	Southeastern Interior Longleaf Pine Woodland	0.00%	0.00%	0.29%	Forest/Woodlands
36	Southern Appalachian Low Mountain Pine Forest	0.00%	0.00%	8.70%	Forest/Woodlands
37	Southern Coastal Plain Blackwater River Floodplain Forest	3.61%	5.56%	0.00%	Floodplain forest
38	Southern Coastal Plain Nonriverine Cypress Dome	0.00%	0.13%	0.00%	Floodplain forest
39	Southern Interior Acid Cliff	0.00%	0.00%	0.00%	other
40	Southern Interior Calcareous Cliff	0.00%	0.00%	0.00%	other
41	Southern Interior Low Plateau Dry-Mesic Oak Forest	0.00%	0.00%	0.00%	Forest/Woodlands
42	Southern Piedmont Cliff	0.00%	0.00%	0.00%	Other
43	Southern Piedmont Dry Oak-(Pine) Forest - Hardwood Modifier	0.00%	0.00%	0.64%	Forest/Woodlands
44	Southern Piedmont Dry Oak-(Pine) Forest - Loblolly Pine Modifier	0.00%	0.00%	0.08%	Forest/Woodlands
45	Southern Piedmont Dry Oak-(Pine) Forest - Mixed Modifier	0.00%	0.00%	0.09%	Forest/Woodlands
46	Southern Piedmont Mesic Forest	0.00%	0.00%	0.11%	Forest/Woodlands
47	Southern Piedmont Small Floodplain and Riparian Forest	0.00%	0.00%	0.04%	Floodplain forest
48	Southern Ridge and Valley Dry Calcareous Forest	0.00%	0.00%	20.74%	Forest/Woodlands
49	Southern Ridge and Valley Dry Calcareous Forest - Pine modifier	0.00%	0.00%	1.32%	Forest/Woodlands
50	Unconsolidated Shore	0.01%	0.01%	0.00%	other
51	Undifferentiated Barren Land	0.06%	0.07%	0.64%	Other