STRATEGIES TO IMPROVE TAPER AND HEIGHT ESTIMATES IN SOUTHERN PINES GROWING IN THE SOUTHEAST UNITED STATES

By

ALLISON BELLE SHEEKS

(Under Direction of Cristian R. Montes)

ABSTRACT

Forest inventories aim at recording tree attributes, which require large amounts of time in the field and are not exempt of errors. These are combined to estimate volume and product yield. Research has shown that including more than one diameter along the stem improves total and merchantable yield estimates. This could be accomplished using remotely sensed light detection and ranging (LiDAR) evaluated from the ground. Unfortunately, LiDAR values often underestimate total height. This research proposes methods to determine tree diameter for loblolly pine using circle fitting to LiDAR data, a tree's total height by solving a taper equation for total tree height and taper simultaneously for longleaf pine, and the effect of water deficit as a determinant of stem taper for longleaf pine plantations in Georgia. The results show a clear relation of locally estimated Kozak parameters with water stress, highlighting the need to include this variable into taper studies.

INDEX WORDS: longleaf pine, LiDAR, *Pinus palustris*, taper equation, water deficit

STRATEGIES TO IMPROVE TAPER AND HEIGHT ESTIMATES IN SOUTHERN PINES GROWING IN THE SOUTHEAST UNITED STATES

By

ALLISON BELLE SHEEKS

B.S., University of Georgia, 2019

A Thesis submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2021

STRATEGIES TO IMPROVE TAPER AND HEIGHT ESTIMATES IN SOUTHERN PINES GROWING IN THE SOUTHEAST UNITED STATES

By

ALLISON BELLE SHEEKS

Major Professor: Cristian R. Montes

Committee: Bronson P. Bullock
Joseph Dahlen

Electronic Version Approved:

Ron Walcott
Dean of the Graduate School
The University of Georgia
December 2021

ACKNOWLEDGEMENTS

I would like to thank my family for their unwavering support and encouragement throughout this entire process. There was not one moment where I had to wonder if they believed I would succeed at the University of Georgia in undergrad or graduate school.

All of this was made possible with the funding from the Warnell School of Forestry, Natural Resources Conservation Service, and the Plantation Management Research Cooperative. Not only did this help financially, but it also provided me with teaching and presentation experience, which I will be able to use beyond the scope of my education.

Morgan Sandera and Brendan Calhoun have consistently reminded me that I am exactly where I need to be. Laura Ramirez, Mark Porter, Thomas Harris, and Stephen Kinane have offered much sage wisdom about the entire Master's process and what happens after. I am especially grateful to Laura as she always made time to help explain a topic that I was struggling with, or even to just be a sounding board for new ideas

Perhaps the most important person I want to thank is Dr. Cristian Montes. He served as my undergraduate advisor, and is 90% of the reason why I decided to pursue a Master's degree. There are so many great professors at Warnell, but Cristian is one of, if not the best. He sincerely wants all his students to understand what they're learning, instead of just memorizing information. This trait has made this entire process much less daunting. I never once thought to myself "is this a dumb question?" because I knew he would happily explain that concept as many times as I needed until I fully understood it. I have never met another professor quite like him, and I don't imagine I ever will.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS……iv

LIST OF FIGURES…….....vii

LIST OF TABLES……..…….…….……....x

CHAPTER

1 INTRODUCTION AND LITERATURE REVIEW……...1

1.1 INTRODUCTION…….…..1

1.2 LITERATURE REVIEW……3

1.3 RATIONALE AND SIGNIFICANCE……6

1.4 GOALS AND OBJECTIVES……..6

1.5 METHODS.…….6

1.6 REFERENCES……8

2 COMPARING DIAMETERS ESTIMATED WITH LIDAR TO DIAMETERS TAKEN IN THE FIELD…….……...11

2.1 INTRODUCTION…….…….…….....11

2.2 METHODS…….……...11

2.3 RESULTS……..15

2.4 DISCUSSION AND CONCLUSIONS……..……...….16

2.5 REFERENCES……..19

2.6 TABLES AND FIGURES…….……...……..…21

3 USING TAPER MODELS TO ESTIMATE TOTAL HEIGHT……..…….23

3.1 INTRODUCTION…….……23

3.2 METHODS…….……...…24

3.3 RESULTS…….…….....29

3.4 DISCUSSION AND CONCLUSIONS…….30

3.5 REFERENCES……..33

3.6 FIGURES…….…….…….……...……35

4 LOCALIZING STEM TAPER EQUATIONS USING WATER DEFICIT

IN LONGLEAF PINE…….……..38

ABSTRACT…….…….38

4.1 INTRODUCTION…….……....39

4.2 METHODS…….…….……..41

4.3 RESULTS……..46

4.4 DISCUSSION AND CONCLUSIONS…….……...….47

4.5 REFERENCES…….…….…….…….……..49

4.6 TABLES AND FIGURES…….……...…….…….51

LIST OF FIGURES

2.1 Cross section of a measured tree from a study site in Virginia. Each small circle is a point collected with the LiDAR scanner, and the larger circular outline is the tree ring estimation from the diameter extraction equation…….…….…….…….…….……...21

3.1 Plot of residuals and predicted relative diameters squared from the Max and Burkhart (1976) when using all diameter measurements……..35

3.2 Plot of residuals and predicted relative diameter squared from the Max and Burkhart (1976) when using only diameter measurements taken below 6 meters……...35

3.3 Relative height and relative diameter squared for 400 trees in South Georgia. The data is fit using the Max and Burkhart model with felled tree measurements, no estimated total heights……36

3.4 Relative height and relative diameter squared for 400 trees in South Georgia. The data is fit using the Max and Burkhart model with felled tree measurements taken up to six meters, where λj is the estimated total height per tree j…….36

4.1 Map of twenty study sites location in Georgia, United States, where each red dot represents one site.……..……...51

4.2 A histogram of the variance of the B0 values estimated from the Kozak taper model for Georgia, United States…….……..52

4.3 A histogram of the variance of the B1 values estimated from the Kozak taper model for Georgia, United States……...53

4.4 A histogram of the variance of the B_2 values estimated from the Kozak taper model for Georgia, United States…..…….54

4.5 Plot of the B_0 taper estimates from the Kozak taper model for Georgia, United States, where the colors correspond to the different parameter values. The gray bars on the top and right hand side of the plot show the trends in parameter values from left to right, and from top to bottom, respectively……55

4.6 Plot of the B_1 taper estimates from the Kozak taper model for Georgia, United States, where the colors correspond to the different parameter values. The gray bars on the top and right hand side of the plot show the trends in parameter values from left to right, and from top to bottom, respectively……56

4.7 Plot of the B_2 taper estimates from the Kozak taper model for Georgia, United States, where the colors correspond to the different parameter values. The gray bars on the top and right hand side of the plot show the trends in parameter values from left to right, and from top to bottom, respectively……57

4.8 Plot of the B_0 taper estimate uncertainty from the Kozak model for Georgia, United States, where the colors correspond to the uncertainty. The gray bars on the top and right hand side of the plot show the trends in uncertainty from left to right, and from top to bottom, respectively……...58

4.9 Plot of the B_1 taper estimate uncertainty from the Kozak model for the Georgia, United States, where the colors correspond to the uncertainty. The gray bars on the top and right

hand side of the plot show the trends in uncertainty from left to right, and from top to bottom, respectively……...59

4.10 Plot of the B_2 taper estimate uncertainty from the Kozak model for the Georgia, United States, where the colors correspond to the uncertainty. The gray bars on the top and right hand side of the plot show the trends in uncertainty from left to right, and from top to bottom, respectively……...60

4.11 Map of water deficit (in millimeters) for Georgia, United States……..61

LIST OF TABLES

2.1 Diameter estimation for the 69 trees used to fit the regression model to predict remaining twenty percent of the trees, where Act. DBH is the measured DBH in the field, Est. DBH is the estimated diameter obtained from LiDAR from the fitting process, difference is the difference between the actual and estimated DBH, and n is the number of trees measured from each stand……..22

2.2 Diameter predictions from the regression model described in the previous table, where Act. DBH is the measured DBH in the field, Est. DBH is the estimation from the fitting process in Chapter 4, difference is the difference between the actual and estimated DBH, and n is the number of trees from each stand…….....22

4.1 Stand description for the twenty sites used in the longleaf pine study. All sites were located in South Georgia……62

4.2 Summary statistics for the taper models tested……..……63

4.3 Minimum and maximum values for the parameters calculated using the Kozak taper model……..……63

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

This thesis begins with a chapter including a brief introduction to longleaf pine (Pinus palustris Mill.) that illustrates and highlights its importance for the methods developed as part of this research. The concept of tree taper as one of the key models that permit volume and yield calculations is explained, as well as the rationale and significance behind this project.

1.1 INTRODUCTION

Longleaf pine has had a tumultuous presence in the United States. Prior to European colonization, longleaf pine occupied approximately 92 million acres (Frost 1993). After decades of exploitation, destruction from turpentine operations, and land clearing for growing cotton, the longleaf pine acreage in the United States dwindled significantly (Walker et al. 2006). By 1997, longleaf pine forest land was reduced to 0.4 million acres (New Georgia Encyclopedia 2019). In 2010, a longleaf pine initiative was established in order to reestablish longleaf pine to various states that were once part of its native range (NRCS 2013). As a result of this significant conservation effort, there are currently 4.3 million acres across the southern United States, spanning from Louisiana to North Carolina, most of which is planted (Oswalt et al. 2012). Longleaf pine is more resistant to disease, effects of natural disasters, and fire than loblolly pine and slash pine (Oswalt et al. 2012). Longleaf pine can grow straighter, and its wood is stronger than loblolly pine (Oswalt et al. 2012) making it an attractive species for pole production. The

longleaf pine ecosystem supports a number of unique plant and animal species. According to Walker (1993), the longleaf pine ecosystem supports 187 plant species that are rare or vulnerable to extinction. Longleaf pine stands also provide habitat for 29 threatened and endangered wildlife species (NRCS 2013). In addition to providing environmental benefits, longleaf pine can also provide financial benefits in the form of pine straw, land for hunting leases, and timber (Sunday 2012). The potential financial return is dependent on stem volume, and an accurate estimate is vital to determining the purchase and selling price for forest landowners. Taper equations can be used to estimate stem volume. There are currently several taper equations in existence, the most popular of which utilize the diameter at breast height (DBH) and the total height of a tree. Equations such as the Kozak (Kozak et al. 1969), modified Kozak, Ormerod (1973), and the Max and Burkhart (1976) have been used. The first three are simple taper models, and the latter is a segmented taper function. Each model uses a ratio of height and diameter measurements. Taper equations that utilize more diameters along the stem could give a more accurate measurement of the taper because there are more intervals of diameter decrease along the stem (Sabatia 2015).

Measuring diameters and total heights in the field is already time consuming but increasing the number of diameters would further exacerbate that issue. A potential solution to that problem is ground-based light detection and ranging (LiDAR). LiDAR units can take full scans of several trees in seconds, but requires an increase in processing time, which was performed using R Software (R Core Team, 2018). LiDAR scans will provide an essentially limitless number of diameters in a fraction of the time, which is invaluable when estimating taper. The required amount of labor in the field will also be decreased, as it only takes one person to hold the unit and walk through the woods. The unit is simple to use, so no formal or extensive training would be required to use it. The analysis in R can also be done by one person. For a

small plot of land, the change may not seem substantial, but for large swaths of land with hundreds, thousands, or millions of acres, this would be extremely valuable.

1.2 LITERATURE REVIEW

Accurate stem volume estimates are essential in determining the potential and current value of a forested stand. Being able to measure this in a timely matter is critical, especially when selling or buying land. The Kozak taper model (1969) was integrated to calculate stem volume. This volume equation only required the use of DBH and total height (Ht) to estimate stem volume. This was problematic considering the differences in form and taper among and between different tree species. Kozak modified his taper equation to include parameters that ensured that the diameter at the top of the tree would equal zero regardless of the other values used in his equations. Ormerod (1973) added a condition to his equation that ensured that the predicted diameter at 4.5 ft. would match the measured DBH. Sharma and Oderwald (2001) combined these conditions in an equation to make sure that both conditions (top diameter = 0, measured DBH matched predicted DBH) were met. However, they found that the differences in predictions were hardly varied and were not worth the extra effort. Max and Burkhart (1976) created a segmented taper equation and it combines three different functions with two different join points. This equation was then integrated and is now the most common method to estimate stem volume. Max and Burkhart's equation is widely used, as it is flexible enough to use for all species of pine, despite having markedly different growth patterns.

Diameter measurements collected via LiDAR can save time and money, granted their measurements are comparable to hand-measured data. Henning and Radtke (2006) completed a project using loblolly pine and found that the diameter measurements gathered using LiDAR were 1—2 cm different from the caliper measurements. Though it is slightly less accurate than

the caliper measurements, the ease of calculating subsequent diameter measurements farther up the stem can be useful when calculating stem taper and estimating stem volume.

The addition of environmental variables, such as water deficit could improve upon the taper estimations. Longleaf pine is long regarded as a drought tolerant species, and according to Samuelson and Stokes (2019), it can be an ideal species to plant as a forest management strategy to combat climate change. Samuelson et al. (2019) conducted an experiment where they simulated drought by creating troughs that would catch the throughfall each time it rained. The DBH, basal area, water use efficiency, and mortality were not significantly affected, however, the volume decreased by 21% in relation to the control group. The growth efficiency of trees subjected to the drought treatments also suffered from a 32% decrease in growth efficiency. A decrease in volume growth over time would influence taper and could be useful in predicting stem taper measurements. Additionally, a study conducted by Kidombo and Dean (2018) found that defoliation or a decreased leaf area significantly reduced the stem and branch growth in 4-year-old loblolly pine.

Since the relationship between environmental variable and stem taper is unknown, using standard parametric functions may not be ideal. If the wrong relationship form is chosen, there is a good chance that the estimated values will have a high error. There are several types of nonparametric regression including kernel estimators, splines and wavelets (Faraway 2006). The type used in Chapter 4 is a type of spline called a thin plate spline (TPS), which was originally developed by Duchon (1975). The main goal of TPS is to minimize the residual sum of squares of a function. However, this minimization is subject to a roughness penalty. If it is too smooth it will not be a good fit of the data, if it is too rough then it is fitting almost all of the points in the data set. Fitting all the points would decrease the predicting capabilities for points outside of the

data set. It can also be used to see the relationships between several variables. Unfortunately, splines can suffer the "curse of dimensionality," which means that as the number of variables increases, the sample size will drastically increase. If there is not enough data, splines will not be the best option (Faraway 2006).

To solve this dilemma, additive models can be used, and were developed by Stone (1985). These models can easily plotted and the relationships between the variables can visually be seen. Additive models use the same base as linear models, but each of the predictor variables will now have some kind of smoothing function attached to it. The parameters are also fit simultaneously as opposed to one by one in linear models. This will be useful when looking to visualize if water deficit has an effect on stem taper.

Forest landowners need to be able to estimate the amount of available wood on their properties. The current taper equations only utilize the DBH and total height of the tree, which can lead to overestimations or underestimations. On smaller stands, these variances may seem inconsequential, but on tracts of lands that cover hundreds of acres, these inaccurate predictions can cost forest landowners a significant amount of financial loss. Georgia produces the nation's largest amount of wood, and therefore, could benefit from the creation of a better stem volume estimation model. Currently, no studies have created a localized taper model for any pine species. The longleaf pine initiative is trying to increase the number of longleaf pine ecosystems in Georgia and across the U.S. Studies showing the amount of volume produced in longleaf pine forests in relation to loblolly pine could persuade forest landowners to transition their plantations from loblolly pine to longleaf pine. Also, in lieu of traditional cruising expenditures, annual savings in companies can be reinvested for areas of higher return.

1.3 RATIONAL AND SIGNIFICANCE

There is currently little data on taper and volume estimations for longleaf pine planted at relatively wide spacing in the state, and the National Resources Conservation Services (NRCS) is curious to know what the volume and growth rates in comparison to other species. If ground-based LiDAR is proven to be successful at capturing the measurements of trees in Georgia, then the same methods can be used for longleaf pine plantations in other states and regions. Eventually, the equations can be fit to other species of pine as well. Using LiDAR will significantly reduce the time and energy needed to do field inventories on stands.

1.4 GOALS AND OBJECTIVES

The goal of the project is to improve stem volume estimations of standing timber for forest landowners by accurately estimating stem taper. The first objective is to compare the data collected via destructive sampling and the data collected via ground-based LiDAR to ensure that LiDAR can be used to capture the diameter measurements along the stem of the tree. The second objective is to estimate the total height of trees using the best taper model. The third objective is to add a localized parameter to the taper model with the best fit.

1.5 METHODS

The data for this research was collected across South Georgia as well as in Appomattox, Virginia.

1.5.1 SITES

The data for Chapter 2 was collected in Virginia in cooperation with Dr. Corey Green at Virginia Tech. There are eight different pre-established loblolly pine stands utilized for this

chapter. There are three plots located in each of the stands, which correspond to different levels of thinning treatments, which include a control, an intermediate thin, and a heavy thin. Dr. Green selected five trees from each of the plots, for a total of 120 sampled trees.

The data for Chapters 3 and 4 were collected across the southern half of Georgia. There were approximately 20 longleaf pine sites; 11 were old-field sites and 9 were cutover sites; this aspect was not used as part of any of the analysis. More details about the specific aspects that factored into site selection can be found in Harris (2020).

1.5.2 DATA

The data was collected at twenty different sites, eleven of which were old-field sites and the remainder were cutover sites. The trees were measured standing and once they were felled. Individual diameter measurements were taken at various heights using a logger's tape. The longitude and latitude pints were taken at each site. Approximately 20 trees were measured and felled at each site, for a total of 400 trees. More information about tree selection and site selection can be found in Thomas Harris' thesis (2020). All trees and their corresponding measurements were compiled into a tree dataset that is used for Chapters 3 and 4. Some edits were made for easier access to information. The measurements were converted to metric units and subsisted to only include data where dob^2/DBH^2 is less than 1.2, where dob refers to the diameter outside the bark. There were several form issues when this number exceeded 1.2. This implies the dob are over 9 percent larger than the DBH. Logically this should only happen if there is tree swelling or large knob in the tree. Chapter 2 uses a database of loblolly pine tree data collected via LiDAR scanning in Virginia via Dr. Corey Green form Virginia Tech. Some trees were repeated, or did not have a recent diameter measurement included, so they were removed from the dataset.

1.6 REFERENCES

Duchon, J., 1975. Linear estimation of non-stationary spatial phenomena. In: Advanced Geostatistics in the Mining Industry. Reidel, Dordrecht, pp. 49-68.

Faraway, J. 2006. Extending the Linear model with R. Boca Raton, Fla: Chapman & Hall/CRC p 234-245, 251

Frost, C. 1993. Four Centuries of Changing Landscape Patterns in the Longleaf Pine Ecosystem. *Proceedings of the 18th Tall Timbers Fire Ecology Conference*, Tallahassee, 30 May-2 June 1991, 299-303.

Harris, T. 2020. Taper, volume, and green weight equations for planted longleaf pine in Georgia, United States. (Master's thesis, University of Georgia, Athens, Georgia).

Henning, J. , and Radtke, P. 2006. Detailed Stem Measurements of Standing Trees from Ground-Based Scanning Lidar. *Forest Science*. 52(1): 67–80.

Kidombo, S., and Dean, T. Growth of tree diameter and stem taper as affected by reduced leaf area on selected branch whorls. *Canadian Journal of Forest Research*. 48(4): 317-323

Kozak, A., Munro, D., and Smith, J. 1969. Taper Functions and Their Application in Forest Inventory. *The Forestry Chronicle* 45(4): 278–83.

Lovell, J.L.., Jupp, D., Newnham, G., and Culvenor, D. 2011. Measuring Tree Stem Diameters Using Intensity Profiles from Ground-Based Scanning Lidar from a Fixed Viewpoint. *ISPRS Journal of Photogrammetry and Remote Sensing* 66(1):46–55.

Max, T. A., and. Burkhart, H. 1976. Segmented Polynomial Regression Applied to Taper Equations. *Forest Science*. 22(3):283–89.

Natural Resources Conservation Service Georgia. 2013. Longleaf Pine Initiative.

Ormerod, D. W. 1973. A Simple Bole Model. *The Forestry Chronicle* 49(3): 136–38.

Oswalt, C. M., J. A. Cooper, D. G. Brockway, H. W. Brooks, J. L. Walker, K. F. Connor, S. N. Oswalt, and R. C. Conner. 2012. History and Current Condition of Longleaf Pine in the Southern United States. Gen. Tech. Rep. SRS–166. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station. 51 p. 166 2012: 1–51

R Core Team. 2018. *R: A Language and Environment for Statistical Computing.* R: Foundation

Sabatia, C and Burkhart, H.. 2015. On the Use of Upper Stem Diameters to Localize a Segmented Taper Equation to New Trees. *Forest Science*. 61(3):411–423.

Samuelson, L., Stokes, T., Ramirez, M., and Mendonca, C. 2019. Drought tolerance of a *Pinus palustris* plantation. *Forest Ecology and Managemen.t* 45.

Sharma, M and Oderwald, R. 2001. Dimensionally Compatible Volume and Taper Equations. *Canadian Journal of Forest Research* 31 (5):797–803.

Stone, C. 1985. Additive regression and other nonparametric models. *Annals of Statistics 13*, 689-705

Sunday, J. 2012. A Financial Look at Longleaf Pine. Georgia Forestry Commission.

Walker, J. 1993. Rare Vasculare Plant Taxa Associated with the Longleaf Pine Ecosystems: Patterns in Taxonomy and Ecology.

Walker, L. R., Jose, S., Jokela, E. J., and Howarth, R.W. 2006.*The Longleaf pine ecosystem: ecology, silviculture, and restoration*. Springer Science & Business Media pp. 429.

CHAPTER 2

COMPARING DIAMETERS ESTIMATED WITH LIDAR TO DIAMETERS TAKEN IN THE FIELD

2.1 INTRODUCTION

Ground-based light detection and ranging (LiDAR) can successfully capture scans of individual trees as well as entire stands than can be viewed easily in statistical software such as R (R Core Team 2018). However, LiDAR data files must be converted and processed to extract usable stand and tree measurements. Ground-based LiDAR is a useful option when several diameters measurements up the stem are needed or if many trees need to be sampled, especially if the trees cannot be destructively sampled. Previous work (Porter, 2021) shows that fitting a circle to cross sectional LiDAR data has led to a good approximation of diameters. The approach yields potentially cleaner data than the one presented before by Fang and Strimbu (2017) who estimated diameters out of the outer most pixel values in two directions only. This chapter will discuss how diameters can be extracted from las files created from LiDAR scans as well as different methods for adjusting the diameters estimations.

2.2 METHODS

The trees used in this data set were previously measured by a team at Virginia Tech University. Each site was managed with three different thinning intensities including a control, a moderately managed site, and an intensively managed site. The diameters were measured and

recorded at breast height. An excel file with the stand ID number, tree tag, plot number, DBH, total height, and crown height were provided for the different site locations.

In November 2020, 120 trees were individually scanned using a Zeb Horizon LiDAR unit, which has a 100m range (GeoSLAM Ltd., Nottingham, UK). A large metal stake was placed at an angle in front of the desired tree to ensure tree identification when looking at the LiDAR scans later on the computer. Each stand had fifteen trees measured, with five from each thinning intensity level. The scans were saved from the LiDAR data logger onto a flash drive, uploaded into the GeoSlam Hub (GeoSLAM Ltd., Nottingham, UK) and converted to las files. The GeoSLAM hub software allows the selection of number of points to be extracted from the raw files. After careful inspection of the point cloud with R Software (R Core Team, 2018), nine percent of the points provides an acceptable picture of the stand. However, using fewer points than the ones scanned, limited the possibility to discern a circular diameter shape per each individual tree or even the general outline of the tree. Including 100 percent of the points provides the clearest image of the tree and circular diameter, but the time need to convert to a las file takes five minutes per tree on average instead of the one minute for the nine percent default of points.

In addition to an increase in the amount of time to convert to a las file, the computations in R are more challenging. There are several steps to get the las files into a usable form for diameter extractions. First the ground layer must be classified. This is necessary to ensure that the height of 0 ft. is correctly classified. If it is not classified first, there is no way to guarantee that each tree is being measured at the same height. The trees were located on sites with various slopes, so what may be 0 ft. for one tree may be on a hill 30 feet higher for a different tree.

Next, the tree heights must be normalized. This portion takes the longest and is main reason why the percentage of points must be adjusted. The normalization process only takes three minutes per tree when using nine percent of the data, however, the scans were not usable as there were not enough points to estimate diameters. Using 100 percent of the points is not feasible; it takes 31 minutes on average to normalize a single tree. For 120 trees, this would take almost three days if the code was run nonstop, and even longer if there were more trees, also subject to human error. If there were 1000 trees, it would take over seventeen days. To mitigate this issue, different percentages were tested, timed, and the tree scans were compared. The distinct circular cross-sections of the trees were visible up to six meters when using 100 percent of the points. Scans with 80% of the points were similarly visible. Any lower than that, and the tree rings were less distinct, or not distinct at all. There were also four trees that were unable to be normalized in R.

After each individual tree was normalized, they were saved as a new las file to ensure they would not need to be normalized again. Each scan included several trees, so the scans had to be clipped to include the individual tree. Three clipping methods were tried, the first was clipping in CloudCompare (CloudCompare 2.11.3, 2020). However, this proved to be difficult as the understory was very thick, and the tree identification stake was hardly visible. The next attempt was to plot the points and isolate the spot with the highest concentration of points. Each scanned tree was circled twice with the LiDAR scanner and should have the highest concentration of points, unfortunately, this was not always the case. The scanner was held at eye level, so the heights of the trees were not fully recorded. There were other smaller understory or midstory trees that ended up having a higher concentration of points. Because of this, the third method was employed; manual tree selection. This involved plotting each tree in R and

subsetting each las file to the area that had the tree in question. This was still an expedient process and took less than a minute to clip a tree. Despite a quick clipping time, there were a few small issues. Some trees were covered by too much understory, so it was not possible to locate the desired tree. After going through these issues there were 89 trees left.

After the 89 trees were isolated, the diameters were extracted with each tree being analyzed separately. A function was created in R to define the radius and the diameter of the circle using the equation of a circle:

$$r^2 = (x-h)^2+(y-k)^2 \quad (2.1)$$

where:

x = x coordinate of circle point

h = x coordinate of center point

y=y coordinate of circle point

k = y coordinate of center point

r = radius of circle

This function was designed to be flexible to allow for a variable center point. This is helpful if the cross section at the height in question is not a perfect circle, as the diameter for most trees is not a perfect circle. To compare the diameters at breast height, the elevation, represented by Z, was defined as a range between 1.37 meters and 1.375 meters. A range was used instead of the exact height measurement to allow for the best results. The starting points were the average X and Y coordinates. A plot with the points from the LiDAR scan, the proposed center of the circle, and a line that traced shape of the circle as determined by the

function were created, and an example of what this looks like can be found in Figure 2.1 The estimated diameters were compared to the hand measured measurements; however, the average difference was -462.4727 centimeters, which is not possible. After inspecting the data further, trees 772 and 534 had diameters that were impossibly high diameters and were removed from the data set. After their removal, estimations improved. Two other alternatives were tested to try and further improve diameter estimations. Instead of using a range between 1.37 and 1.375, diameters between 1.38 and 1.385 and 1.36 and 1.365 were estimated and the two were averaged together to create a set of new averaged diameters. The new set were compared with the measured diameters. The second alternative was to use those same new values to fit eighty percent of the trees from with linear regression and then predict the diameters for the remaining twenty percent. The trees were randomly selected to be used for regression or to be predicted. This would create an adjustment factor for future LiDAR scans.

2.3 RESULTS

After the removal of tree 772 and 534, the average diameter difference between the measured diameter and the LiDAR estimated diameter was 3.02 centimeters. This was still a little high, so the two other alternatives were tested. The new average diameters were compared to Dr. Green's and yielded an average difference of 2.95 centimeters. This is an improvement from using the measurements from 1.37 meters to 1.375 meters only. Fitting a regression and predicting diameters yielded the best results; the average difference between points when using linear regression was only 0.05 centimeters. Additionally, the R^2 value for the regression is approximately 0.87. See Table 2.1 for all RMSE and bias values for each method.

The averages by stand for the data used to fit the linear regression model can be found in Table 2.2. The averages by stand for the predicted values for the remaining 20% of the trees can

be found in Table 2.3. The diameter predictions start to over predict diameters around 25 centimeters and higher with an average 1.40 centimeter difference. Trees with diameters of less than 25 centimeters tend to under predict diameters with an average 0.8 centimeter difference.

2.4 DISCUSSION AND CONCLUSIONS

The LiDAR unit can create a good general image of a stand and trees, however, upon further inspection the number of points above six meters in height are severely lacking. This makes accurate diameter estimation almost impossible above that height. Perhaps the scans could be improved by angling the scanner up towards the top of the trees, or even holding the scanner above one's head when collecting the scans.

Ideally, the target tree would be located in the middle of the scan and it would have the highest concentration of points. This is very easy to identify when processing in R. However, many midstory trees had a higher concentration of points, so the tree cropping could not be automated and had to be done manually. Before that process can begin, the scans must first be normalized to get the trees in the form needed to extract diameters. It took the CPU roughly three days of consistent working to finish the 87 trees used in the sample. If the sample size increased to 1000 or even just 200 trees, the processing time would increase drastically. Until the vertical height scanning is improved, automating the dimeter extraction process is not possible.

Also, there 43 trees were rendered unusable due to thick understory or an inability to be normalized. This is roughly one third of the total trees in this sample. This is a large reduction in the sample size and could have an effect of the accuracy of diameter estimations. Using linear regression to create an adjustment factor was helpful to get more accurate diameter estimation, however there were only 18 trees used in the validation data set, and 69 were used to fit the

regression model. To confidently say that these values will accurately estimate diameters in this area of Virginia, let alone the entire state, more trees will need to be sampled.

Additionally, the diameter estimations were off by approximately three centimeters each time, which can be the difference in DBH classes. Fitting the data via linear regression greatly improved this, decreasing the difference six fold. Using these adjustment factors should be the preferred method. However, there are negative implications with that conclusion as well; without the set of diameters measured in the field, the LiDAR estimates would have been skewed. Furthermore, the measured diameters would need to be collected at around the same time the LiDAR scans are taken, or the data fitting will either overestimate or underestimate the diameters again. If the same adjustment factors for this chapter were used for a different part of Virginia or a different state, the diameter estimations could still be quite inaccurate, which would make hand measuring the diameters a safer option in that case.

A study conducted estimated the DBH of a variety of trees along a street as well as in a park using Stencil laser scanning system (Heo et al. 2019). The root mean squared errors (RMSE) were 3.77 cm and 8.95 cm. along the road and park, respectively. These errors are quite high, but the cross sections of the stems were irregularly shaped and the fitting methods were for circular cross sections (Heo et al. 2019). Huang et al. (2011) used a terrestrial LiDAR unit to estimate DBH of Liaotung oak (*Quercus liaotungensis* Koidz.) and black locust (*Robinia pseudoacacia* L) using a Riegl LMS-360i which has a 360° horizontal scanning range and 90° vertical scanning range. The data fitting was successful and they were able to determine DBH with a RMSE of 3.4 cm. (Huang et al. 2011) It is important to note that the trees measured in both of these studies were not any species of pine (*Pinus spp.*) and instead were various hardwood species.

A different model for fitting diameters could be tested to see if improvements on diameter estimations could be made. A study by Koreň et al. (2017) had an absolute value of only 1 cm or less for most of their multi-scan circle fitting methods for European beech trees (*Fagus sylvatica* L.) Perhaps using one of these methods would improve the prediction in this chapter. Another study conducted by Brolly and Király (2009) used terrestrial laser scanning to estimating DBH for sessile oak (*Quercus petraea*), hornbeam (*Carpinus betulus*), beech (*Fagus sylvatica* L.), larch (*Larix decidua*) and spruce (*Picea abies*) species. They found that their models underestimated the DBH significantly with lower bias values, but relatively high RMSE values.

After examining different estimation methods, using ground-based LiDAR and processing the data in R is be the best option with the lowest bias and RMSE values. Despite being the best option, there are still several aspects that can be improved before it can fully replace the current methods.

2.5 REFERENCES

Brolly, G. and Géza, K. 2009. Algorithms for Stem Mapping by Means of Terrestrial Laser Scanning. *Acta Silvatica et Lignaria Hungarica*. 5. 119-130.

CloudCompare (version 2.11.3) [GPL software]. (2020). Retrieved from http://www.cloudcompare.org/

Fang, R. and Strimbu, B.M. 2017.Stem Measurements and Taper Modeling Using Photogrammetric Point Clouds. *Remote Sensing*. 9(7):716

Heo, H., Lee, D., Park, J., and Thorne, J. 2019. Estimating the heights and diameters at breast height of trees in an urban park and along a street using mobile LiDAR. *Landscape Ecol Eng* **15,** 253–263.

Huang, H., Li, Z., Gong, P., Cheng, X., Clinton, N., Cao, C., Ni, W., and Wang, L. 2011. Automated Methods for Measuring DBH and Tree Heights with a Commercial Scanning Lidar. *Photogrammetric Engineering and Remote Sensing*. 77. 219-227.

Hub GeoSLAM. 2020. Available online: https://geoslam.com/solutions/geoslam-hub/

Koreň, Milan, Mokroš, Martin, Bucha, Tomáš. 2017. Accuracy of tree diameter estimation from terrestrial laser scanning by circle-fitting methods. *International Journal of Applied Earth Observation and Geoinformation*, Volume 63, 2017, Pages 122-128.

Porter, M. 2021. Diameter and height estimation using a lightweight handheld terrestrial laser scanner in longleaf pine plantations. (Master's thesis, University of Georgia, Athens, Georgia).

R Core Team 2018. *R: A Language and Environment for Statistical Computing.* R: Foundation for Statistical Computing, Vienna, Austria.

GeoSLAM. 2020. Zeb Horizon: User's Manual. Nottingham, UK: GeoSLAM Ltd.

2.6 TABLES AND FIGURES

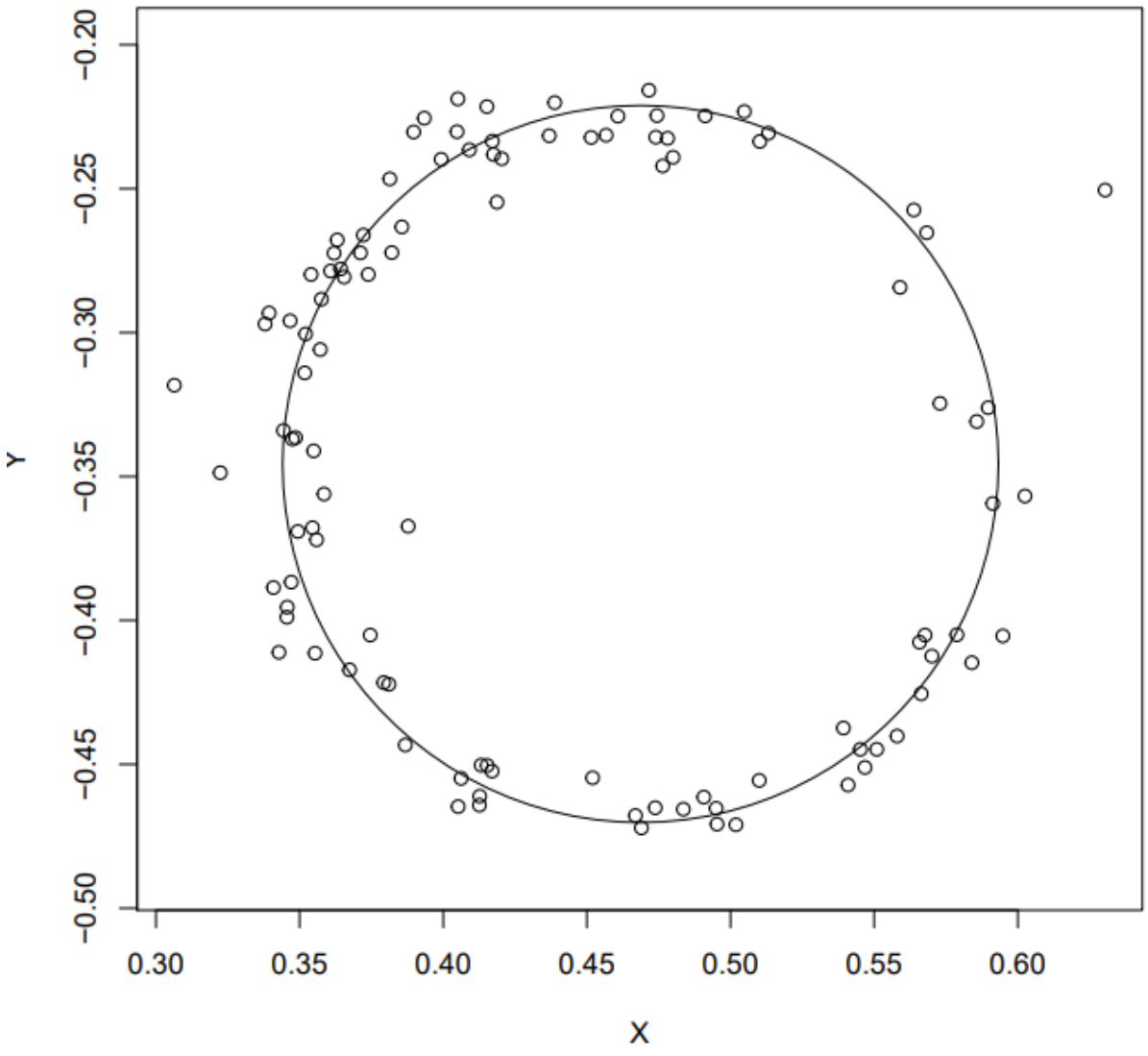

Figure 2.1: Cross section of a measured tree from a study site in Virginia. Each small circle is a point collected with the LiDAR scanner, and the larger circular outline is the tree ring estimation from the diameter extraction equation.

Table 2.1: Diameter estimation for the 69 trees used to fit the regression model to predict remaining twenty percent of the trees, where Act. DBH is the measured DBH in the field, Est. DBH is the estimated diameter obtained from LiDAR from the fitting process, difference is the difference between the actual and estimated DBH, and n is the number of trees measured from each stand.

Stand #	Act. DBH (cm)	Est. DBH (cm)	Difference (cm)	n
1301	31.0	34.59	-0.36	8
1303	26.1	28.84	0.54	9
1304	30.2	33.65	1.15	8
1306	29.1	31.99	-1.04	7
1307	30.1	33.55	0.45	7
1308	28.1	31.20	-1.57	9
1309	30.5	33.85	0.06	12
1314	26.6	29.41	0.37	9

Table 2.2: Diameter predictions from the regression model described in the previous table, where Act. DBH is the measured DBH in the field, Est. DBH is the estimation from the fitting process in Chapter 4, difference is the difference between the actual and estimated DBH, and n is the number of trees from each stand.

Stand #	Act. DBH (cm)	Est. DBH (cm)	Difference (cm)	n
1301	29.2	32.56	0.97	2
1303	23.71	25.72	1.46	1
1304	29.8	33.13	-1.38	6
1306	26.8	29.53	0.82	4
1307	29.4	33.45	-2.97	1
1308	-	-	-	0
1309	26.2	27.47	-1.06	2
1314	30.1	35.34	-1.43	2

CHAPTER 3

USING TAPER MODELS TO ESTIMATE TOTAL HEIGHT

3.1. INTRODUCTION

Accurately measuring total height in the field is the hardest aspect of field inventory, as well as the most expensive. A common method of measuring tree height is via hypsometer, a hand held tool that relies on geometrical relations to determine total height (Avery & Burkhart, 2002). These devices are handy and useful because they only require one person to operate it. However, care must be taken to properly measure the distance between the hypsometer and the target tree to avoid estimation bias error. New hypsometers overcome this limitation using lasers to determine the distance to the target tree, but users should still be at least the distance of one tree height away from the tree that is being measured (Bettinger et al. 2017). There are several environmental factors that can impede them from being able to get one tree height away from the one they are measuring. If the brush is too thick, or if the crowns overlap too much, especially when pine trees can reach upwards of 80-100+ feet. Finding ways to streamline the height measurements will decrease the cost of inventory itself.

Ground based light detection and ranging (LiDAR) can collect a massive amount of points at varying elevations, which could expedite the data collection process. LiDAR can also solve the problem from standard hypsometers any time operators do not have to be one tree

height away. They have less of a limitation with respect to distance to the target tree and the flexibility to allow for any direction with respect to the tree to be used as reference point. Additionally, if the understory is particularly thick or under conditions impeding a clear path of the tree, an operator can simply bypass that issue and will still be able to obtain tree measurements of that tree as well as several others over a certain distance. However, there is still an optical limitation in determining the total tree height from the farthest return on a given tree and that is the impossibility to have a clear line of sight to the tip, particularly under high productive stands with elevated leaf production. To bridge this gap, it is hypothesized that tree form, as determined by a series of diameters measured at fixed heights, will allow the estimation of total height when taper equations is solved for a given tree. This chapter will focus on estimating the total height of a tree using taper models and discuss the potential application of LiDAR.

3.2. METHODS

The diameter outside bark (dob), DBH, and total height measurements were collected via destructive sampling across twenty sites in the southern portion of Georgia, United States.

To estimate tree height, two taper models (Kozak et al., 1969, and the Max and Burkhart, 1976) were fitted to data under the assumption that these models properly describe the shape of a tree when several measurements along the tree height are available. Taper models use known height as part of the independent variables, allowing them to estimate a ratio that is used to fit the model. However, when DBH is known, so are diameter at different heights, total height can be solved out of a system of equations.

To test this premise, two different taper equations were first compared to see which would better explain taper. Then, using the best equation, it was solved for height as a local parameter to determine each tree's estimated height.

The first model corresponds to a simple polynomial equation (Kozak, 1969) expressed as follows:

$$\frac{d_i^2}{D^2} = B_0 + B_1 \frac{h_i}{H} + B_2 \frac{(h_i)^2}{H^2} \qquad (3.1)$$

where:

d_i = diameter outside bark (dob) in centimeters

D = diameter at breast height (DBH) in centimeters

h_i = height at a point along the stem in meters

H = total tree height in meters

The second model used is the Max and Burkhart model (1976):

$$\frac{d_{i,j}^2}{D_j^2} = B_1 \left(\frac{h_{i,j}}{H_j} - 1\right) + B_2 \left(\frac{h_{i,j}^{\ 2}}{H_j^2} - 1\right) + B_3 \left(a_1 - \frac{h_{i,j}}{H_j}\right)^2 I_1 + B_4 \left(a_2 - \frac{h_{i,j}}{H_j}\right)^2 I_2 \qquad (3.2)$$

where:

$d_{i,j}$ = dob at point along stem j in centimeters

D_j = DBH for tree j in centimeters

$h_{i,j}$ = height at point along stem j in meters

H_j = Total height of tree j in meters

To run this code in R, starting values for parameters B_0, B_1, and B_2 must be set. Fortunately, those values were already estimated in Chapter One of this thesis, and were used as the starting values for this chapter.

These models were compared using bias and root mean squared error, and were fit using a five-fold cross validation (Tukey 1986). The trees were randomly split into five groups of eighty trees. The data was fit five separate times. For the first iteration, the trees were fit using all folds except the first one, the second iteration fit the trees using all the folds except the second one, and so on, until all folds were used and excluded. The results were used to compare the two different models and see which one provided the best total height estimation.

The resulting best model, was later calibrated to data replacing the observed H_j by local parameters λ_j, which were estimated as part of the process optimization process. To assign local parameters, a dummy matrix was created using each tree's id number as columns with as many observations as the ones in the table with trees. Using the id number as the dummy variable created a unique numerical value for each tree, which means only one regression model is needed and can be applied to all tree id numbers. If the dummy variable method was not used, a separate regression function would have to be created for each tree and their set of measurements.

To solve for λj, a modified version of the Max and Burkhart equation was used:

$$\frac{d_{i,j}^2}{D_j^2} = B_1\left(\frac{h_{i,j}}{\lambda_j} - 1\right) + B_2\left(\frac{h_{i,j}^2}{\lambda_j^2} - 1\right) + B_3\left(a_1 - \frac{h_{i,j}}{\lambda_j}\right)^2 I_1 + B_4\left(a_2 - \frac{h_{i,j}}{\lambda_j}\right)^2 I_2 \quad (3.3)$$

where:

$d_{i,j}$ = dob at point along stem j in centimeters

D_j = DBH for tree j in centimeters

$h_{i,j}$ = height at point along stem j in meters

λ_j = estimated height of tree j in meters

To solve for the parameters, a maximum likelihood approach was used, that required not only the estimation of all model parameters, but also the estimation of a general variance value. After the regression was run, and the estimated parameter values were calculated, a prediction model was run. This used the same code as the regression model, except the starting values for the prediction model were that of the regression model. The official output from the prediction process were the predicted λ_j values, which are the total heights.

The models were fit using a Maximum Likelihood (ML) approach. The ML approach finds the parameters that make a set of observations more likely to have happened. The ML approach starts by defining a distribution for the mean of the expected values, in this case, a normal distribution was used:

$$y_i \sim N(\mu_i, \sigma^2 I) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}} \tag{3.4}$$

Where μ_i is the model functions *f*(β), which represent the Kozak (1969) and the Max and Burkhart models (1976) in this study, and $\sigma^2 I$ is a square matrix whose main diagonal is equal to σ^2

$$:\sigma^2 I_n = \begin{pmatrix} \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma^2 \end{pmatrix} \tag{3.5}$$

The likelihood *L* of a single observation y_i given our vector of parameters (β, which in this study are B_0, B_1, and B_2) and σ^2 is defined as:

$$\mathcal{L}(y_i|\boldsymbol{x}, \boldsymbol{\beta}, \sigma^2) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(y_i - f(x_i\boldsymbol{\beta}))^2}{2\sigma^2}} \tag{3.6}$$

The likelihood of the whole sample is equal to the product of likelihoods from each individual observation:

$$\mathcal{L}(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(y_i - f(x_i\boldsymbol{\beta}))^2}{2\sigma^2}} \tag{3.7}$$

Taking the log of both sides can express the *L* as the log-likelihood as:

$$\ell(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\beta}, \sigma^2) = \sum_{i=1}^{n} log \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(y_i - f(x_i\boldsymbol{\beta}))^2}{2\sigma^2}} \tag{3.8}$$

To find the set of parameters that maximize the log-likelihood for the data set, it is:

$$argmin\, f(x,\beta) = \sum_{i=1}^{\infty} log(\frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(f(x,\beta)-y_i)^2}{2sigma^2}} \quad (3.9)$$

The heights derived from the taper model were then compared to the ground measured total heights as well as the total heights measured with a hypsometer. To see the potential benefits of using LiDAR, several scans were viewed and plotted in R. Individual trees were selected at different points along the stem to see the tallest discernable height point. Unfortunately, the highest point of visibility for these scans was approximately six meters, not the total height of the tree. To gauge the applicability of LiDAR to estimate total height, the entire fitting process was performed again, using felled-tree measurements up to 6 meters for each tree.

All computing methods in this chapter were performed using R Software (R Core Team, 2018).

3.3 RESULTS

The Max and Burkhart model had a lower bias and RMSE as compared to the Kozak model, which is why it was used for the remainder of the chapter. The Max and Burkhart model's estimated heights were the closest to the felled-tree measurements with a mean difference of 0.0028 meters.

When using all felled-tree measurements, there is a total height bias of 0.11 meters. When using felled-tree measurements taken at less than six meters, there is a total height bias of 6.23 m. Plots of the residuals and predicted values for all felled-tree measurements and for felled tree measurements up to six meters can be found in Figures 3.1 and 3.2. Figure 3.1 has residual

values that are consistently close to zero. This means that the predicted values are very close the actual values. Figure 3.2 has some results that are close to zero, but some are also spread more widely, which implies that the predicted values do not align as closely with the actual values. Figure 3.2 only uses measurements six meters up the stem, so it makes sense that the upper height values are not predicted accurately. In both figures there is a diagonal line near the 1.0 value on the predicted scale. This occurs because typically the final point should match up with the total height.

3.4 DISCUSSION AND CONCLUSIONS

The Max and Burkhart model (1976) yielded the best fit in this chapter. Total height was successfully estimated using diameter and height measurements along the stem. Figure 3.3 shows the comparison between relative height and relative diameter squared. There is a straight horizontal portion where the y-axis is 1. This occurs because the diameter outside bark (dob) should equal the DBH. This occurs at 1.372 meters, or breast height. The plot stops at 1.0 on the x axis because that is the maximum relative height, and at that point, we assume that the tip has a relative diameter of zero. This is not the case for Figure 3.4, which was developed using only the heights at or below six meters. As shown in the figure, some of the predicted total heights are less than their actual heights; this is why the values are greater than one. This means that the estimate total heights are under predicted, which would pose an issue when selling timber because the total volume would be under estimated.

Using a limited number of diameter and height measurements lead to a severe height bias, so using only LiDAR measurements in the future could increase that risk unless additional height and diameter measurements could be obtained above six meters. Forking is another area that could potentially exacerbate the height bias issue. Finding ways to improve the heights that

ground-based LiDAR units can reach will solve most of the issues outlined in this chapter. LiDAR would be able to make even greater strides in improving the field inventory process, volume estimation, and subsequently save money both in inventory and the buying and selling process.

The main perk of using LiDAR is a reduced amount of time spent in the field, but it is met with an increase in post processing time. A simple way to test how significant the increase in processing time is, was to increase the sample size. One hundred additional rows were added and filled with random values; the run time increased, and would continue to do so if more trees were added. However, this could potentially be coded in a different programming language that performs more quickly than R. In addition to adding more data points, the number of folds was also changed which resulted in additional run time for both scenarios. The initial starting values were also changed to see if the values yielded were different; they were not.

Ground-based LiDAR to estimate total height is a relatively new forest inventory method, and there are minimal publications that utilize it. Wei et al. (2014) used ground-based LiDAR to measure total height of standing red cypress trees. The RMSE between the LiDAR heights and the field measured heights was only 0.254, which is higher than the RMSE values for the models used in this chapter. Their LiDAR unit was also set up on a stationary stand instead of being moved around the forest. This helps confirm that walking through the plantation can yield better results. A study was conducted with a mobile LiDAR unit at an urban park and along a street in South Korea (Heo et al. 2019). The average residual was approximately 0.4 meters. Heo et al (2019) noted that the total heights when using LiDAR were underestimated.

There are other methods for estimating total height when LiDAR is not available. For example, a study was conducted in Northern Iran that utilized artificial intelligence (AIN) model

to estimate total heights from what starting data? (Bayat et al. 2020). They had 516 trees with hand measured diameters and total heights to fit their dataset, and they found that its models performed better than standard empirical models, with R^2 values of 0.78 for the AIN models and 0.68 for the regression analysis. However, the root mean squared percent error for the AIN model was 0.79% higher than the regression model (Bayat 2020). A different study used a generalized regression neural network (GRNN) to accurately predict total height with an RMSE% of 6.171% to 6.209% for the Brutian pine trees (*Pinus brutia* Ten.) (Diamantopoulou and Özçelik 2012). However, they also had felled tree measurements for diameter and height. The use of drones and aerial LiDAR are potential other options but are more costly to conduct unless done over a large area (Tilley et al 2004). In addition to being more costly to conduct, the amount of training required far supersedes that of a ground-based unit, making ground-based LiIDAR a more economically feasible option.

3.5 REFERENCES

Avery, T. and Burkhart, H. 2002. Forest measurements: Fifth Edition. p. 154.

Bayat, M., Bettinger, P., Heidari, S., Khalyani, A., Jourgholami, M., and Hamidi, S. 2020. Estimation of Tree Heights in an Uneven-Aged, Mixed Forest in Northern Iran Using Artificial Intelligence and Empirical Models. Forests, 11(3), 324.

Bettinger, P., Izlar, B., Harris, T., Cieszewski, C., Conrad, J., Greene, D., Mech, A., Shelton, J., Siry, J., Kane, M., Merry, K., Baldwin, S., and Smith, J. 2017. Handbook of land and tree measurements. p. 121.

Diamantopoulou, M. and Ozcelik, R. 2012. Evaluation of different modeling approaches for total tree-height estimation in Mediterranean Region of Turkey. *Forest Systems*. 21. 383. 10.5424/fs/2012213-02338.

Heo, H., Lee, D., Park, J., and Thorne, J. 2019. Estimating the heights and diameters at breast height of trees in an urban park and along a street using mobile LiDAR. *Landscape Ecol Eng* 15, 253–263.

Kozak, A., Munro, D., and Smith, J. 1969. Taper functions and their application in forest inventory. *The Forestry Chronicle*, 22(3):283-289.

Max, T. and Burkhart, H. 1976. Segmented polynomial regression applied to taper equations. *The Forestry Chronicle*, 45(4):278-283.

R Core Team 2018. *R: A Language and Environment for Statistical Computing.* R: Foundation for Statistical Computing, Vienna, Austria.

Tilley, B., Munn, I., Evans, D., Parker, R., and Roberts, S. 2004. Cost Considerations of Using LiDAR for Timber Inventory. Southern Forest Economics Workers. 43-50.

Tukey, J. and Jones, L. 1986. The collected works of John W. Tukey Vol. (4) 639.

Wei, C., Chen, C-T, Chen, J-C., and Wu, S-T. 2014. Using ground-based LiDAR data to measure standing trees in a red cypress plantation. *Taiwan Journal of Forest Science*. 29. 169-178.

3.6. FIGURES

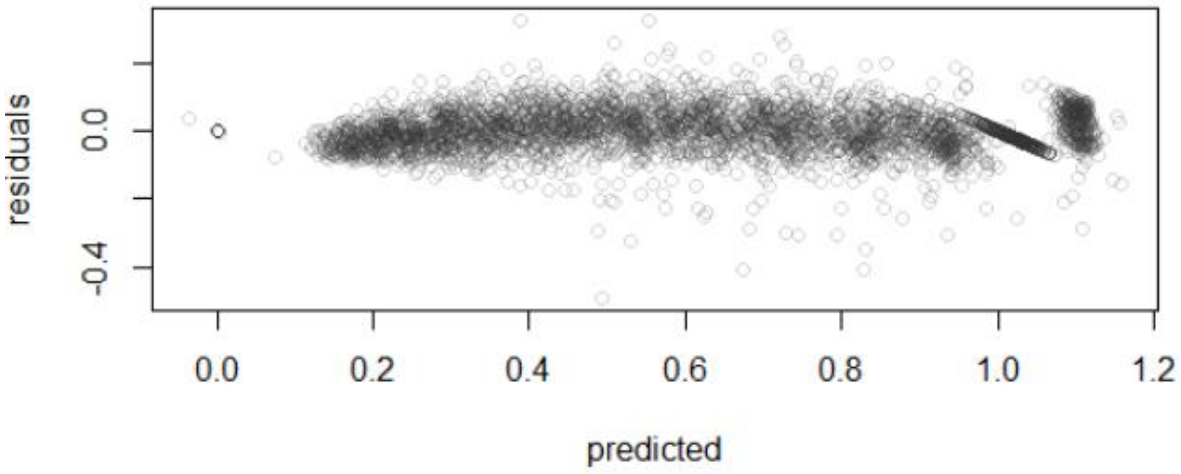

Figure 3.1: Plot of residuals and predicted relative diameters squared from the Max and Burkhart (1976) when using all diameter measurements.

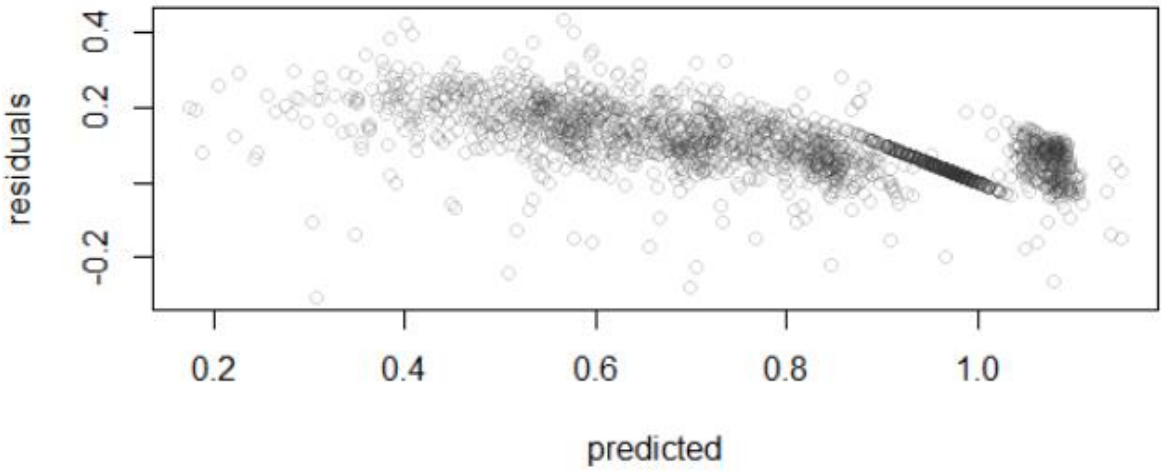

Figure 3.2: Plot of residuals and predicted relative diameter squared from the Max and Burkhart (1976) when using only diameter measurements below 6 meters.

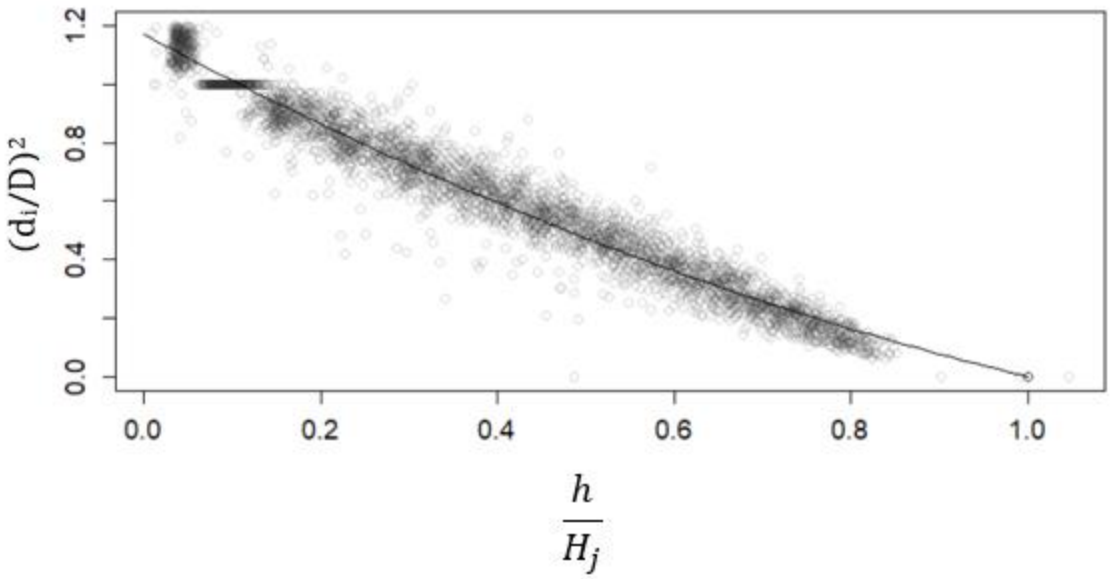

Figure 3.3: Relative height and relative diameter squared for 400 trees in South Georgia. The data is fit using the Max and Burkhart model with felled tree measurements, no estimated total heights.

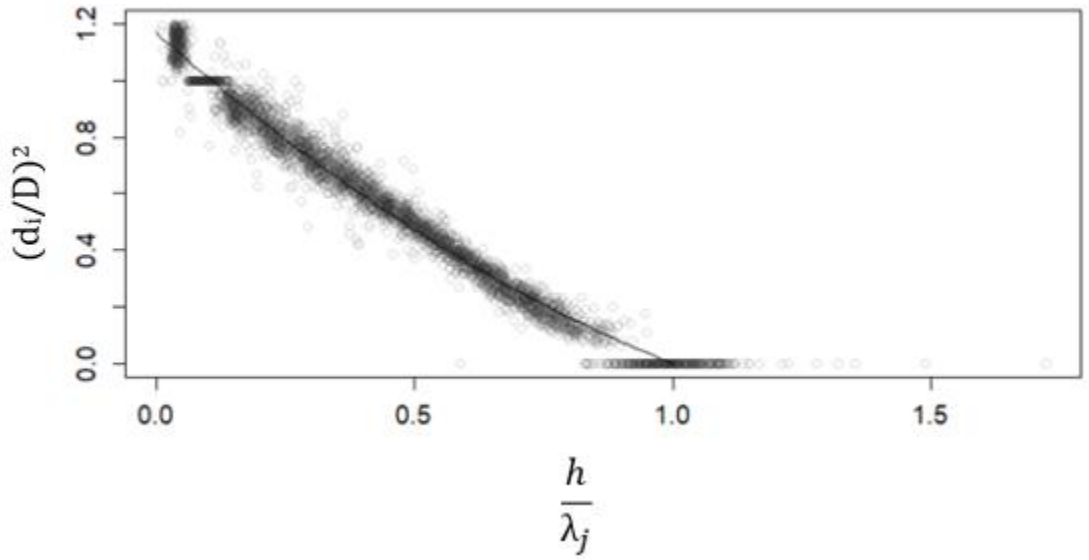

Figure 3.4: Relative height and relative diameter squared for 400 trees in South Georgia. The data is fit using the Max and Burkhart model with felled tree measurements up to six meters, where λj is the estimated total height per tree j.

CHAPTER 4

LOCALIZING STEM TAPER EQUATIONS USING WATER DEFICIT IN LONGLEAF PINE[1]

[1] Sheeks, A., Montes, C., Bullock, B., and Dahlen, J. To be submitted to *Forest Ecology and Management.*

ABSTRACT

The topic of this chapter pertains to the modeling of tree stem's taper and how location and environmental conditions, specifically water deficit, affect stem taper. Four hundred longleaf pine (*Pinus palustris* Mill.) trees were destructively sampled across 20 sites in Georgia, United States. All of the taper data were fit using R (R Core Team 2018) to four different taper models including Kozak (1969), Max and Burkhart (2006), Ormerod (1973) and a modified Kozak equation (Demaerschalk & Kozak 1977). The Kozak model has the lowest bias and is used for the final estimations and predictions. The average B_0, B_1, and B_2 values are 1.298, -2.083 and 0.8067. Water deficit occurs when the potential evapotranspiration exceeds the actual precipitation. This study finds that the estimated B_0 values are higher in areas with a greater water deficit and the B_1 and B_2 values are lower in areas with a greater water deficit.

4.1 INTRODUCTION

The topic of this chapter pertains to the modeling of tree stem's taper and how location and environmental conditions, specifically water deficit, affect stem taper. The stem of the tree is the major perennial part of the tree, and the target of all forest yield studies due to its economic importance. Branches and leaves are considered to be important only for either biomass estimation, specifically branches, or for productivity assessment, specifically foliage. The tree stem is what is commonly purchased and sold in current timber markets. The general form of the tree could be seen as a long cylinder, followed by a cone at the top. However, this would not account for the minor or major changes in taper up the stem. Therefore, trees are described as having three different parts. The largest portion of the tree at the base is neiloid, the middle and more consistent shape is a parabloid, and the top of the tree, where the taper is the highest is assumed to be a conoid (Burkhart and Tomé 2012). Taper equations utilize this assumption to estimate changes from the stump to the top of the stem. As mentioned previously, taper equations typically use the diameter at breast height and the total height of the tree to estimate the taper.

Variation in stem taper will vary by species and growth patterns. A tree's ability to grow is affected by several different factors such as, access to sunlight, nutrients, water availability, slope, soil type, natural and unnatural disturbances, etc. In addition to stem taper, these factors can also affect the tree canopy growth rate. The tree canopy considers the foliar part of the tree and can consist of several layers (Burkhart and Tomé 2012). In particular, longleaf pine has an ovular and open crown (Gilman and Watson 1994). In a study by Bigelow et al. (2020) longleaf pine in Southwest Georgia to test three different treatments to potentially reduce the effects of climate change on the species, which was a resistance, resilience, and transition treatment. The

resistance treatment involved harvesting all trees that were not longleaf pine. The resilience treatment reduced longleaf pine basal area and kept upland and xeric oaks only. Longleaf pine and upland and xeric oaks were thinned, mesic oaks and pines other than longleaf pine were removed, additionally, turkey oak seedlings were planted (Bigelow et al. 2020). A year after their initial tests were conducted, Hurricane Michael passed within 30 km of their study sites, and the trees were subjected to a high level of wind and stress. The study found that longleaf pine suffered the least amount of damage out of all the species. However, Bigelow et al. (2020) mentioned that the trees had a more open canopy before thinning or the removal of the tree species, and thus the trees left after each treatment may have had enough time to be windfirm, and thus more resistant to the effects of the hurricane. A similar study conducted by Johnsen et al. (2009) in Gulfport, Mississippi studied the effects of Hurricane Katrina on stands consisting of loblolly, slash, and longleaf pine. The study sites were directly affected by hurricane Katrina with winds reaching upwards of 140 km/hr. Johnsen et al. (2009) found that longleaf pine had a 7% mortality whereas, loblolly and slash had 14% and 26% mortality. Johnsen et al. (2009) concluded that longleaf pine is more wind resistant than loblolly itself. They similarly noted that the trees in this study appeared to be more windfirm and were not recently thinned Moore et al. (2015) studied results from a series of trials designed to quantify various wood properties of radiata pine (*Pinus radiate* D.), including wood stiffness, and found that it is directly proportional to stand density, which helps explain the conclusion made by Johnsen et al. (2009).

In addition to an increased wind resistance, longleaf pine is considered a more drought resistant species as well. This is because longleaf pine has deeper and wider spreading root systems than loblolly pine or slash pine (Samuelson et al. 2012). Though longleaf pine can withstand more drought stress than other pine species, this study looked to see if water deficit

across the state affected stem taper, and if so, how did it vary across Georgia. Though often used interchangeably, water deficit and drought are not the same things. A drought occurs because of water deficit, but a deficit of water does not always indicate that an area is in a state of drought. A drought is a prolonged period of time with an abnormally low amount of rainfall, whereas water deficit is when there is more evapotranspiration than there is precipitation. This study focused on how water deficit, not drought, affects stem taper

When estimating taper, coefficients from scientific papers published in a region or state are if there is not enough data to estimate taper per that individual's own land. This can be misleading as trees in different parts of a region, or even in the same state, grow differently. This chapter discusses how taper can change based on one's exact location in a state, and how water deficit or excess can affect it.

4.2 METHODS

4.2.1 STUDY SITE

As previously mentioned, the data for this chapter was collected across the southern half of Georgia (See Figure 4.1). Twenty sites were chosen, with eleven old-field sites and nine cutover sites. More information, including the county, latitude, longitude, origin, as well as year of establishment can be found in Table 4.1.

4.2.2 DATA

All 400 destructively sampled trees were used in the taper estimations. Outside bark diameters (dob), diameters at breast height (DBH), tree identification numbers, longitude, latitude, and total height measured on the ground were used. The total heights from the felled trees were used to fit the models. Each tree was measured for dob at 0.152, 0.609, 1.372, and

2.438 m and every 1.219 m above 2.438 m until a 7.62 cm dob was reached. Measurements were taken using a logger's tape.

4.2.3 COMPARING TAPER MODELS

The model fitting, and subsequent analysis was conducted using R Software (R Core Team, 2018). The model that best fit the data was selected after comparing the average bias and the root mean squared error (RMSE):

$$\text{Bias} = (y_i - \hat{y}_i) \tag{4.0.1}$$

$$\text{RMSE} = \frac{(y_i - \hat{y}_i)^2}{n} \tag{4.2}$$

where y_i = observation and $\hat{y}_i$ = the predicted value for the i^{th} observation, n = sample size. These measure the goodness of fit.

4.2.4 TAPER MODELS

Taper models utilize the diameters and heights at different points along the stem to estimate the measurements for the other portions of the tree. One of the earliest models, proposed by Kozak (1969) used a ratio between height measurements and total height, with three parameters to estimate.

$$\frac{d_i^2}{D^2} = B_0 + B_1 \frac{h_i}{H} + B_2 \frac{(h_i)^2}{H^2} \quad (4.3)$$

where:

d_i = dob in centimeters

D = DBH in centimeters

h_i = height at a point along the stem in meters

H = total tree height in meters

Demaerschalk and Kozak (1977) modified equation 4.3:

$$\frac{d_i^2}{D^2} = B_0 + B_1\left(\frac{h_i}{H} - 1\right) + B_2\left(\frac{(h_i)^2}{H^2} - 1\right) \quad (4.4)$$

where:

d_i = dob in centimeters

D = DBH in centimeters

h_i = height at a point along the stem in meters

H = total tree height in meters

Ormerod (1973) created a non-linear taper model that only has one parameter to be estimated.

$$d_i = D\left(\frac{(H-h_i)}{(H-4.5)}\right)^{B_1} \quad (4.5)$$

where:

d_i = dob in centimeters

D = DBH in centimeters

h_i = height at a point along the stem in meters

H = total tree height in meters

Max and Burkhart (2006) created a more complex model that has been fit to several tree species.

$$\frac{d_i^2}{D^2} = B_1 - 1(Z_i-1) + B_2\left(Z_i^2-1\right) + B_3(a_1-Z_i)^2 I_1 + B_4(a_2-Z_i)^2 I_2 \quad (4.6)$$

where:

$$I_j = \begin{cases} 1 & \text{if } Z \le a_j \\ 0 & \text{if } Z > a_j \end{cases} \quad j=1,2; Z_i = \frac{h_i}{H}$$

After choosing the best model, the variance and spatial correlation were examined. Once confident that there was in fact some correlation, each parameter was estimated using linear regression.

To evaluate the effect of water deficit on taper models, a model that summarizes water balance was used. Water deficit occurs when the potential evapotranspiration (PET) exceeds

precipitation. Conversely, water excess occurs when there is more precipitation than there is evapotranspiration. The equation for water deficit is defined as follows:

$$WD= \sum_{i=1}^{12}\left[E_{t0,i}\geq R_i\right]\left(E_{t0,i}-R_i\right), \quad (4.7)$$

where $E_{t0,i}$ is the potential evapotranspiration (PET) on month *i* as defined by the Hargreaves and Samani (1982) formula, and R_i is the total monthly rainfall over month *i*. To ensure that water deficit and excess are strictly positive numbers, when precipitation is greater than PET, PET is subtracted from the precipitation. When precipitation is less than PET, it is subtracted form PET.

The Hargreaves and Samani (1982) equation uses average radiation and temperature to calculate PET:

$$E_{t0}=0.0023*R_A*(T+17.8)*(T_R)^{0.5} \quad (4.8)$$

where:

R_A = mean radiation (mm/day)

T_R = mean daily maximum temperature – mean daily minimum temperatures (Celsius)

T = mean air temperature (Celsius)

Due to the limitations in the database, it was not possible to fit local parameters at each site. Therefore, the regression had to be constrained to include only sites where local regression convergence was granted. The data was fit using the thin plate spline (TPS) function, utilizing the longitude, latitude, and the water deficit values for the US as the independent variables and the parameter itself as the dependent variable. The regression used a reciprocal of the parameter variance as a weighting factor. The regression was then interpolated, and maps were developed for the whole state. Additionally, a separate analysis was conducted to evaluate the spatial

uncertainty over the whole state. The purpose of this was to graphically display how good the taper parameter estimates would be for the rest of Georgia. Having a tangible graph of uncertainty made it significantly easier to see how good or bad the beta value estimations were.

4.3 RESULTS

Out of the four models, the Kozak model fitting yielded the smallest average bias of -1.46e-17, with the Max and Burkhart model the next closest, with an average bias of -1.77e-06, which is significantly higher than the Kozak model (Table 4.2). The Ormerod model had the poorest fit with the highest bias and RMSE value.

There was some degree of variance between the parameters and the spatial correlation, but it did not appear to be significant, as shown in Figures 4.2 through 4.4. In Figures 4.5 through 4.7, there appears to be a discernable relationship between the beta estimates and geographic locations. The sizes of the circles in these figures are weighted by the variance; the larger the circle, the higher level of variance. Initially, the circles were too small to discern using the default values, so a multiplication factor was applied to each. To ensure that accurate comparisons were made, each of the plots were multiplied by the same factor and their plots have the same dimensions.

The uncertainty of the estimates based on location can be found in Figures 4.8 through 4.10. Site 14, located in Charlton County, had the highest uncertainty with a value of 0, 0.1, and 0.08, for B_0, B_1, and B_2, respectively. Site 13, located in Stewart County also had a high uncertainty with a value of 0.8 for both the B_1 and B_2 values. The other sites had uncertainty levels at or around zero for all parameters. However, when looking at Figures 4.8-4.10, it is clear that the B_1, and B_2 estimations had the most uncertainty, with the southwestern and southeastern

borders being the greatest. It is important to note that the majority of the data points were located in the central portion of South Georgia, so the estimations around the border would not be quite as accurate as those in middle Georgia.

The average B_0, B_1, and B_2 values was 1.298, -2.083 and 0.8067, respectively. The difference between the minimum and maximum beta values is the greatest for B_1 and B_2, which makes sense considering those had the highest levels of uncertainty around their estimations. The specific range of values can be found in Table 4.3. The average residual value was essentially zero ($-1.201\ e^{-17}$). Stand 13 in Stewart County had an error level of 0.8 for the B_1, and B_2 values.

4.4. DISCUSSION AND CONCLUSIONS

The Kozak model was able to successfully estimate taper parameter values at each of the twenty different locations across the southern half of Georgia. As expected, uncertainty in parameter estimates remained close to the values from local predictions and these increase as the estimations moved further away from the study sites. This map highlights locations that could be included in future samples to increase parameter prediction certainty, with a positive effect in volume estimates.

Water deficit (Figure 4.11) and parameter values were found to be positively or negatively correlated, depending on the parameter number. The B_0 values are higher in areas with a greater water deficit and the B_1 and B_2 values are lower in areas with a greater water deficit. This means that water availability and the B_1 and B_2 values are proportionally related, so areas that have an equal or a surplus amount of rainfall compared to the PET, should have higher parameter values than populations in dryer areas. It is also important to note that an area can still have an excess of water even if it has not rained in quite some time. While water deficit appears

to influence parameter estimates, more testing should be conducted with other sites across the state to ensure it varies because of the deficit and not for a separate environmental condition.

There are currently no papers that discuss how water deficit can affect longleaf pine stem taper. There are very few papers that discuss how other environmental conditions such as, soil type, water deficit, stand origin, etc. can affect stem taper in longleaf pine. There was one study (Shaw et al. 2003) that added crown ratio as an additional variable to improve taper estimated, however the taper model used was not the Kozak model, so the parameters are not truly comparable. Crown ratio is also not an environmental variable like water deficit, but it can be affected similarly. Amateis and Burkhart (1987) as well as other authors have conducted studies using loblolly pine, but not longleaf pine. In short, this is the first instance where water deficit is shown to have a possible impact on stem taper.

Furthermore, this is the first time that the parameters were estimated based upon spatial location. Previously, people have had to use parameters that were for a larger area, state, or even region. If they were able to find values for a more specific area, the odds of those values being for longleaf pine, as opposed to loblolly pine, are slim. Loblolly and longleaf have markedly different growing patterns and require different management strategies. Incorrect parameter values would pose a real problem if forest landowners are using taper to subsequently estimate volume. Having a general idea of the expected standing timber over a landowner's property can be further improved using this method of localized taper values, that will ultimately inform purchase or selling values reducing overall financial risk .

Additionally, conducting the same testing for different species of pine in Georgia would grant future insight into local variations due to climatic effects.

4.5. REFERENCES

Amateis, R. and Burkhart, H. 1987. Tree Volume and Taper of Loblolly Pine Varies by Stand Origin. *Southern Journal of Applied Forestry*. 11. 185-189. 10.1093/sjaf/11.4.185.

Bigelow, S., Looney, C. and Cannon, J. 2020. Hurricane effects on climate-adaptive silviculture treatments to longleaf pine woodland in southwestern Georgia, USA. *Forestry: An International Journal of Forest Research*. 94.

Burkhart, H. and Tomé, M. 2012. Modeling Forest Trees and Stands. 10.1007/978-90-481-3170-9.

Demaerschalk, J. and Kozak, A. 1977. The whole-bole system: a conditioned dual-equation system for precise prediction of tree profiles. *Canadian Journal of Forest Research*. 7:488-497.

Gilman, E. and Watson, D. 1994. Pinus palustris Longleaf pine. USDA, Forest Serv., Fact Sheet ST-469.

Hargreaves, G., Samani, Z., 1982. Estimating potential evapotranspiration. J. Irrig. Drain. Eng. 108, 225–230.

Johnsen, K., Butnor, J., Kush, J., Schmidtling, R., Nelson, C. 2009. Hurricane Katrina winds damaged longleaf pine less than loblolly pine. *South. J. Appl. For.*, Vol. 33(4): 178-181

Kozak, A., Munro, D., and Smith, J. 1969. Taper functions and their application in forest inventory. *The Forestry Chronicle*, 22(3):283-289.

Max, T. and Burkhart, H. 1976. Segmented polynomial regression applied to taper equations. *The Forestry Chronicle*, 45(4):278-283.

Moore, J., Osorio, R., McKinley, R., Lee, J., and Dash, J. 2015. Effects of silviculture and seedlot on radiata pine growth, wood properties and end-product quality.

Ormerod, D. 1973. A simple bole model. *The Forestry Chronicle*, 49(3):136-138.

R Core Team 2018. *R: A Language and Environment for Statistical Computing.* R: Foundation for Statistical Computing, Vienna, Austria.

Samuelson, L., Stokes, T., Johnsen, K. 2012. Ecophysiological comparison of 50-year-old longleaf pine, slash pine and loblolly pine, *Forest Ecology and Management*, Vol. (274):108-115

Shaw, D., Meldahl, R., Kush, J., and Somers, G. 2003. A Tree Taper Model Based on Similar Triangles and Use of Crown Ratio as a Measure of Form in Taper Equations for Longleaf Pine. Gen. Tech. Rep. SRS-66. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. p. 8.

4.6 TABLES AND FIGURES

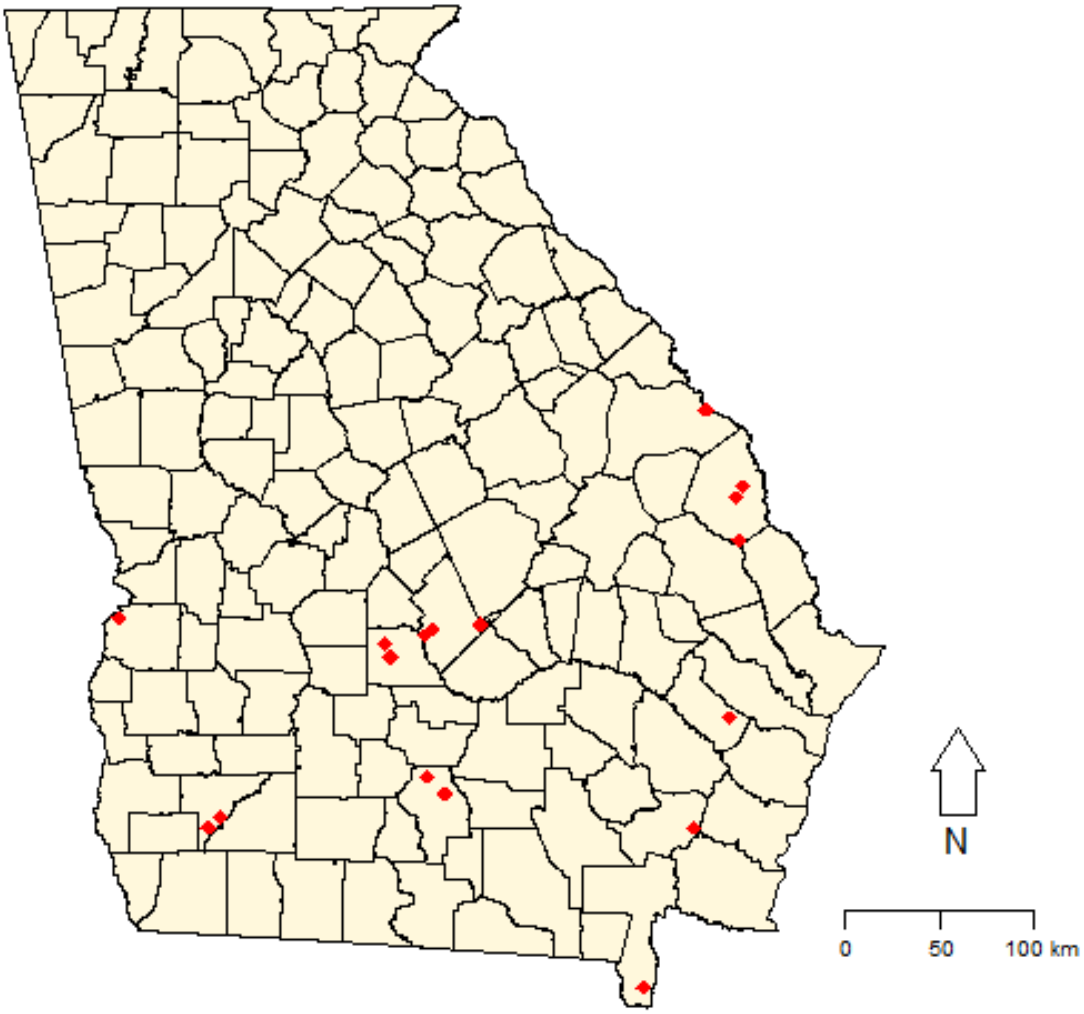

Figure 4.1: Map of twenty study sites location in Georgia, United States, where each red dot represents one site.

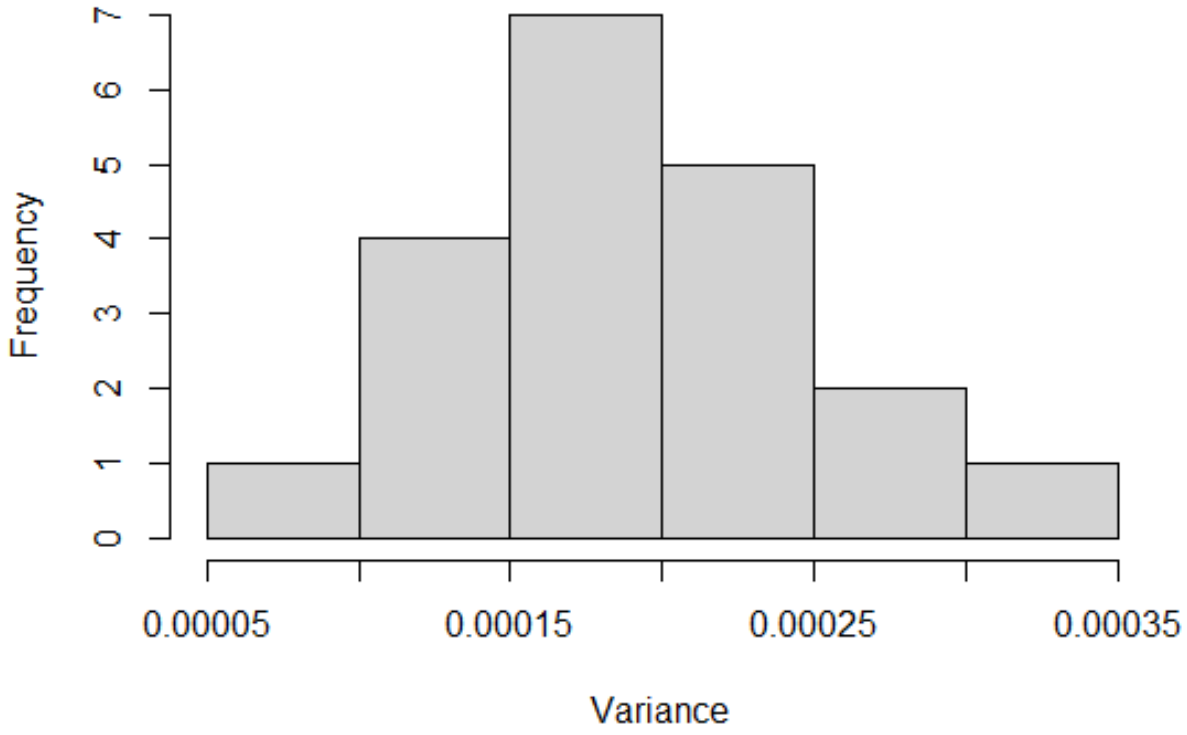

Figure 4.2: A histogram of the variance of the B_0 values estimated from the Kozak taper model for Georgia, United States.

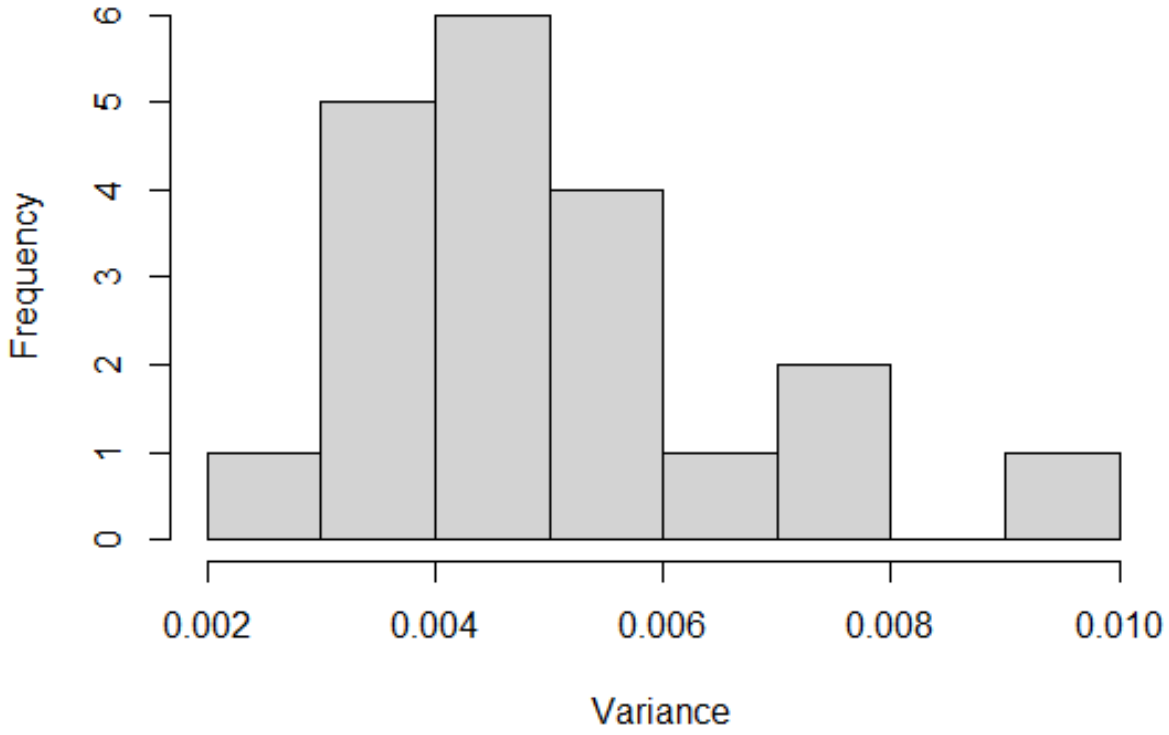

Figure 4.3: A histogram of the variance of the B_1 values estimated from the Kozak taper model for Georgia, United States.

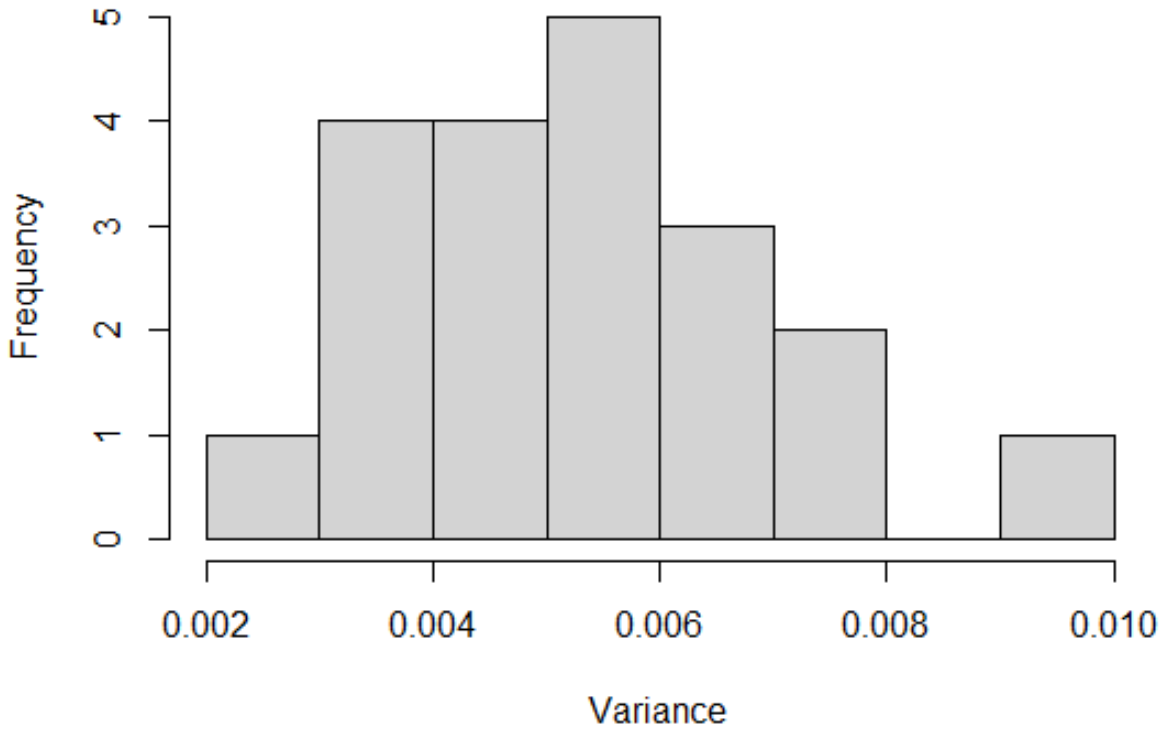

Figure 4.4: A histogram of the variance of the B_2 values estimated from the Kozak taper model for Georgia, United States.

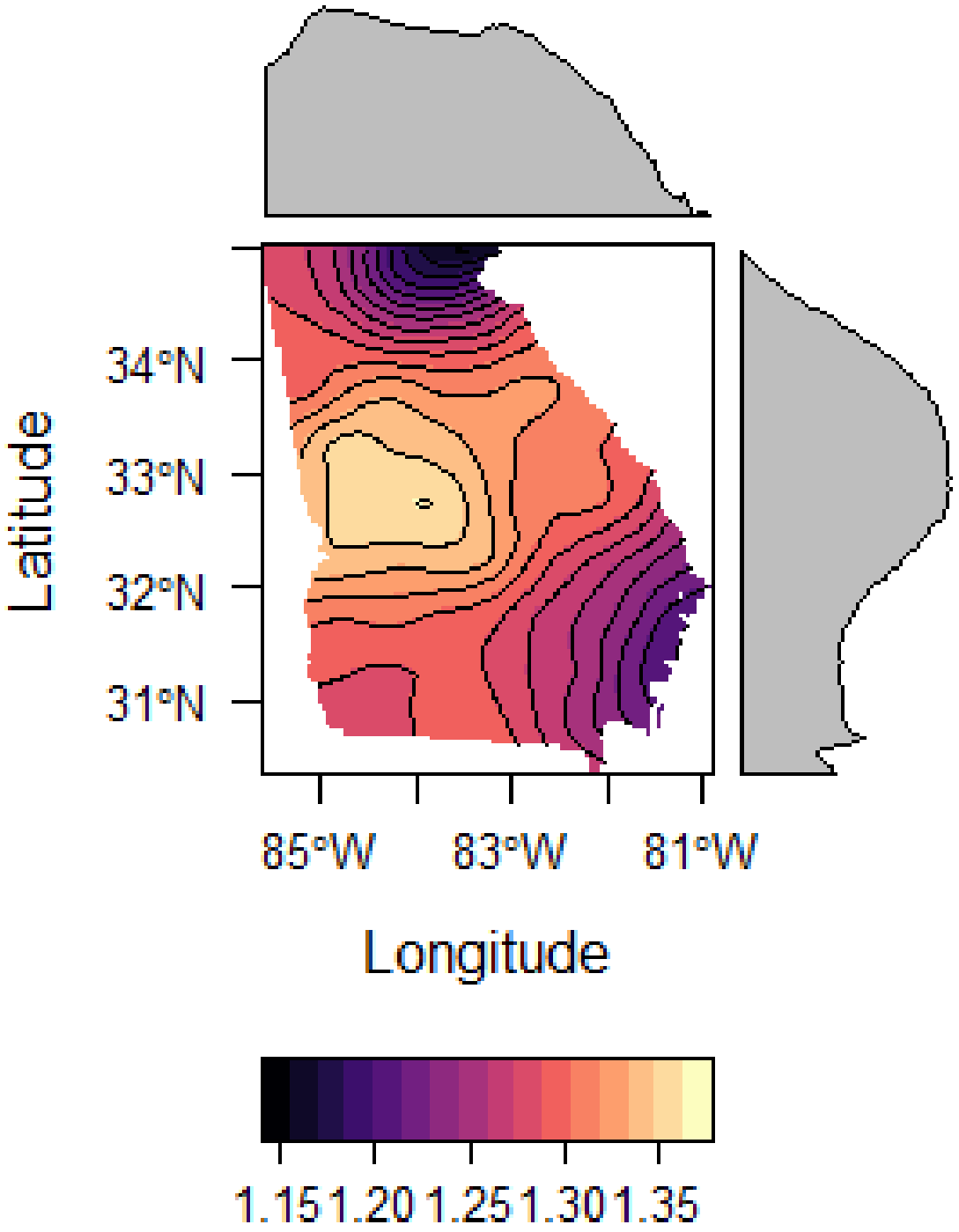

Figure 4.5: Plot of the B_0 taper estimates from the Kozak taper model for Georgia, United States, where the colors correspond to the different parameter values. The gray bars on the top and right hand side of the plot show the trends in parameter values from left to right, and from top to bottom, respectively.

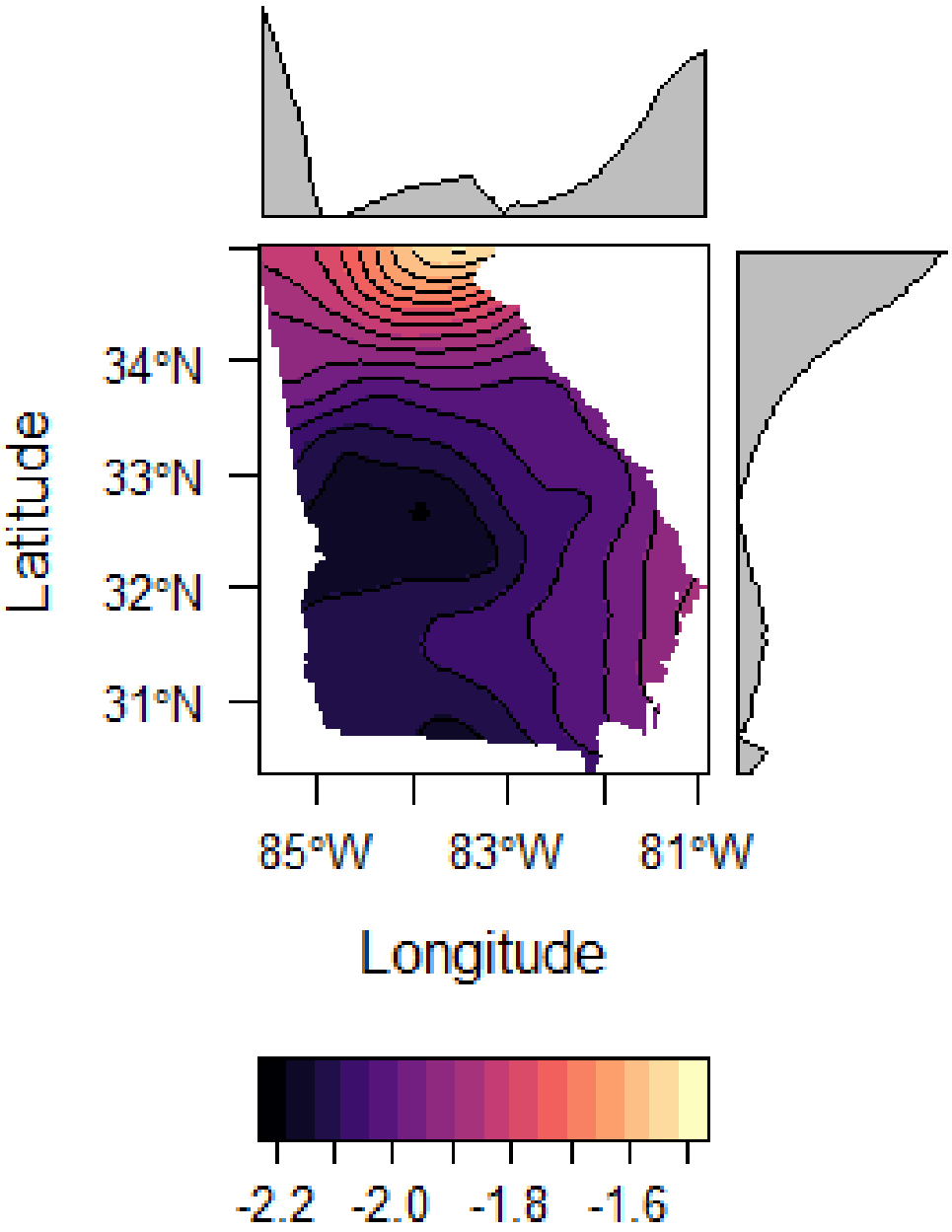

Figure 4.6: Plot of the B_1 taper estimates from the Kozak taper model for Georgia, United States, where the colors correspond to the different parameter values. The gray bars on the top and right hand side of the plot show the trends in parameter values from left to right, and from top to bottom, respectively

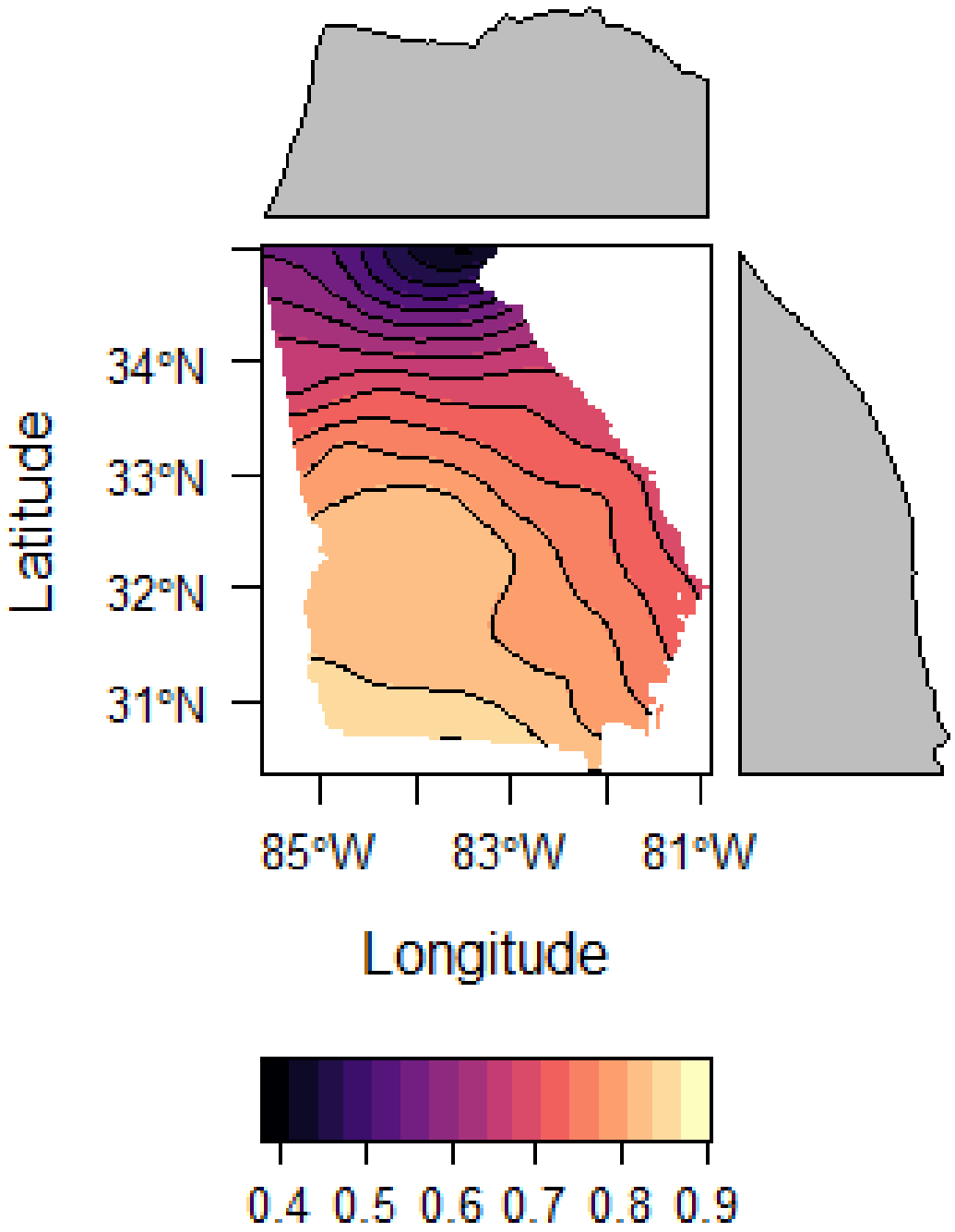

Figure 4.7: Plot of the B_2 taper estimates from the Kozak taper model for Georgia, United States, where the colors correspond to the different parameter values. The gray bars on the top and right hand side of the plot show the trends in parameter values from left to right, and from top to bottom, respectively.

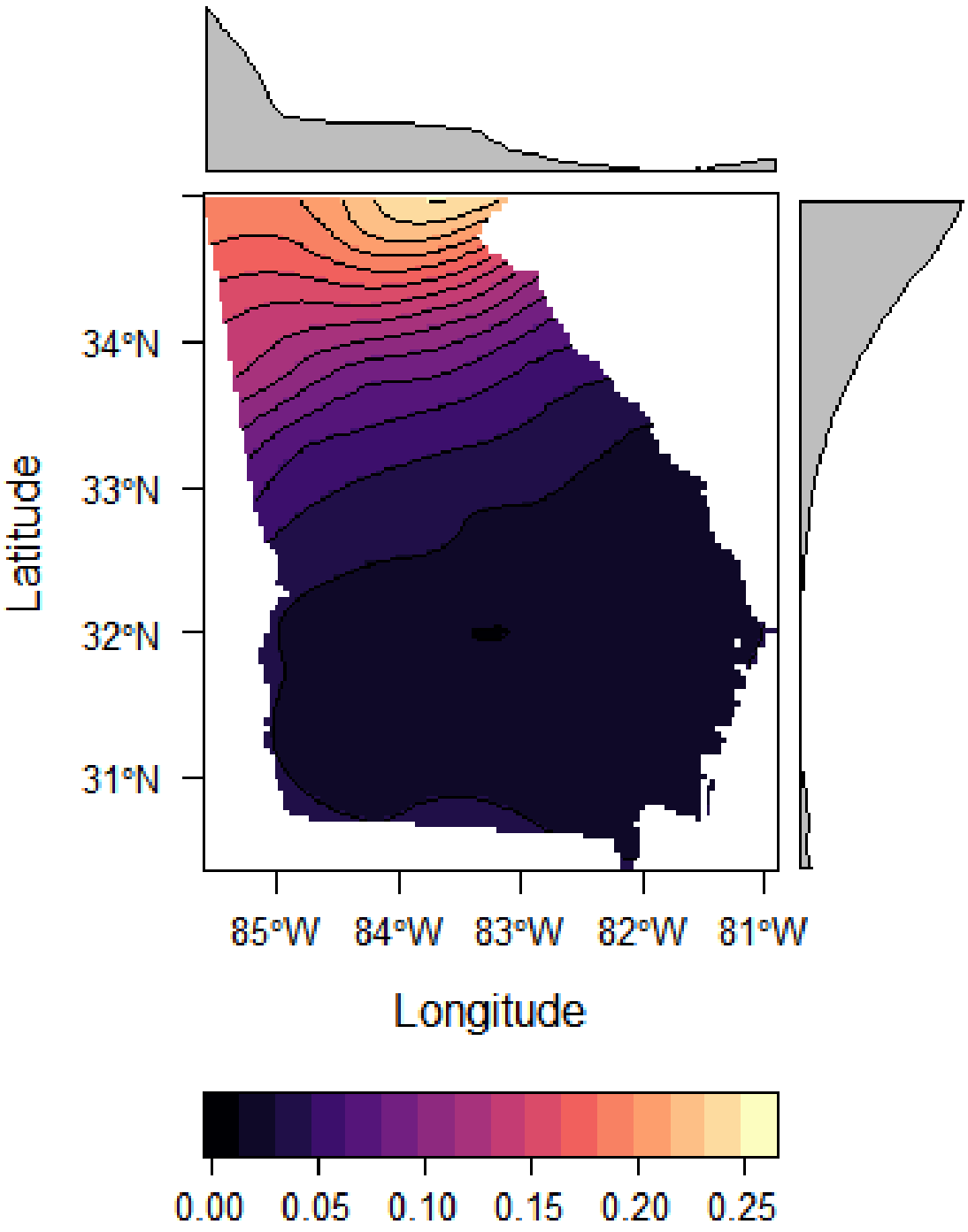

Figure 4.8: Plot of the B_0 taper estimate uncertainty from the Kozak model for Georgia, United States, where the colors correspond to the uncertainty. The gray bars on the top and right hand side of the plot show the trends in uncertainty from left to right, and from top to bottom, respectively.

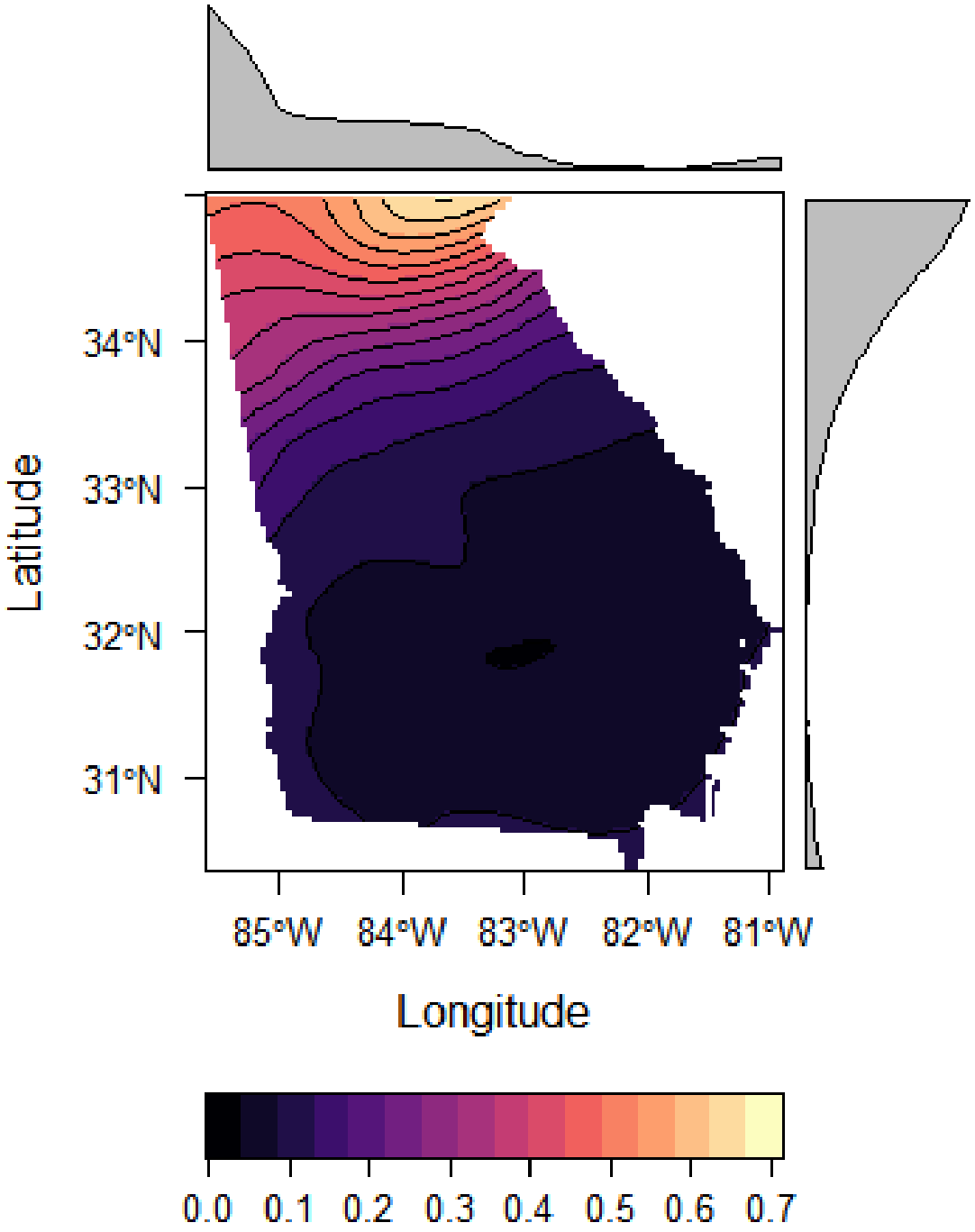

Figure 4.9: Plot of the B_1 taper estimate uncertainty from the Kozak model for the Georgia, United States, where the colors correspond to the uncertainty. The gray bars on the top and right hand side of the plot show the trends in uncertainty from left to right, and from top to bottom, respectively.

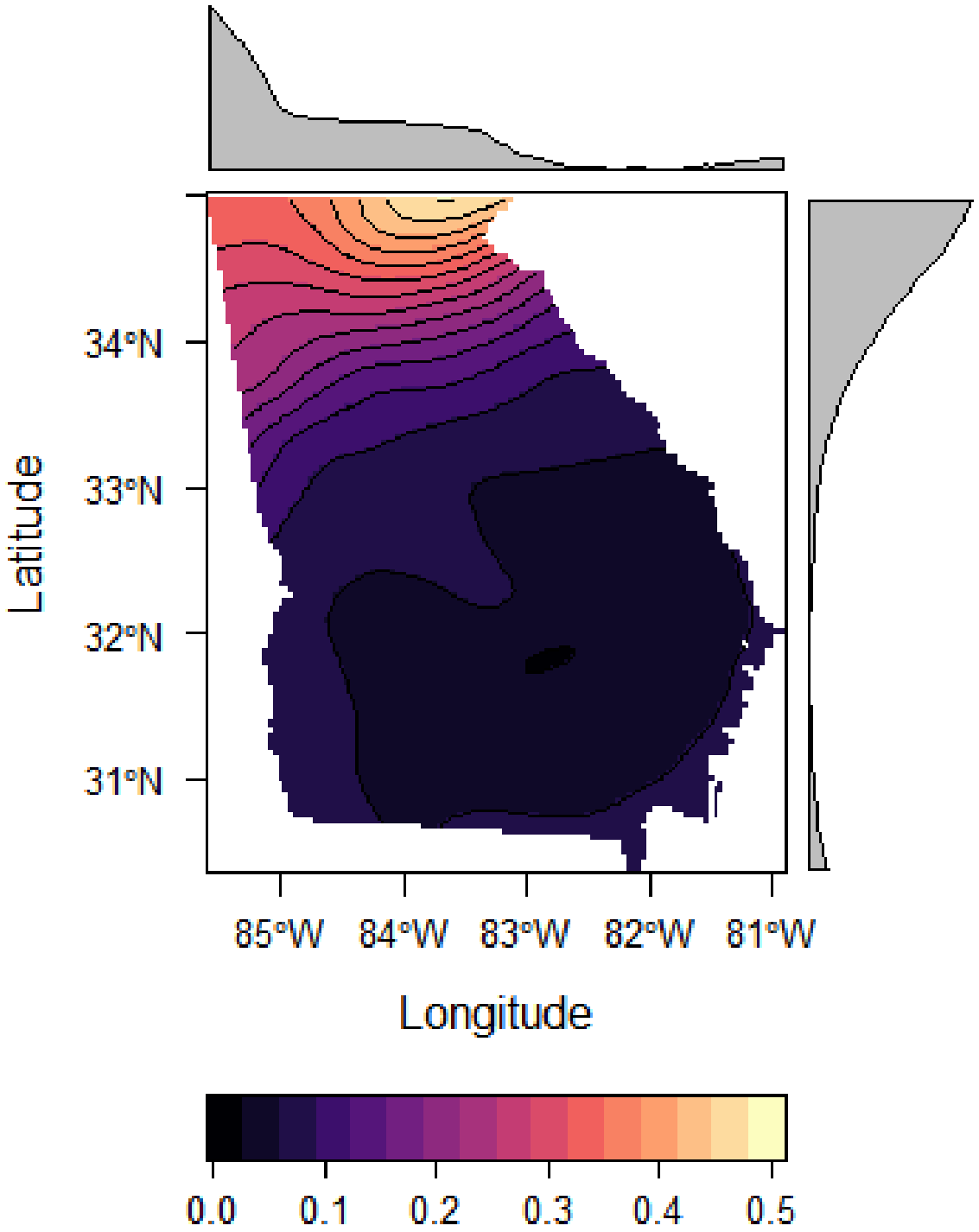

Figure 4.10: Plot of the B_2 taper estimate uncertainty from the Kozak model for the Georgia, United States, where the colors correspond to the uncertainty. The gray bars on the top and right hand side of the plot show the trends in uncertainty from left to right, and from top to bottom, respectively.

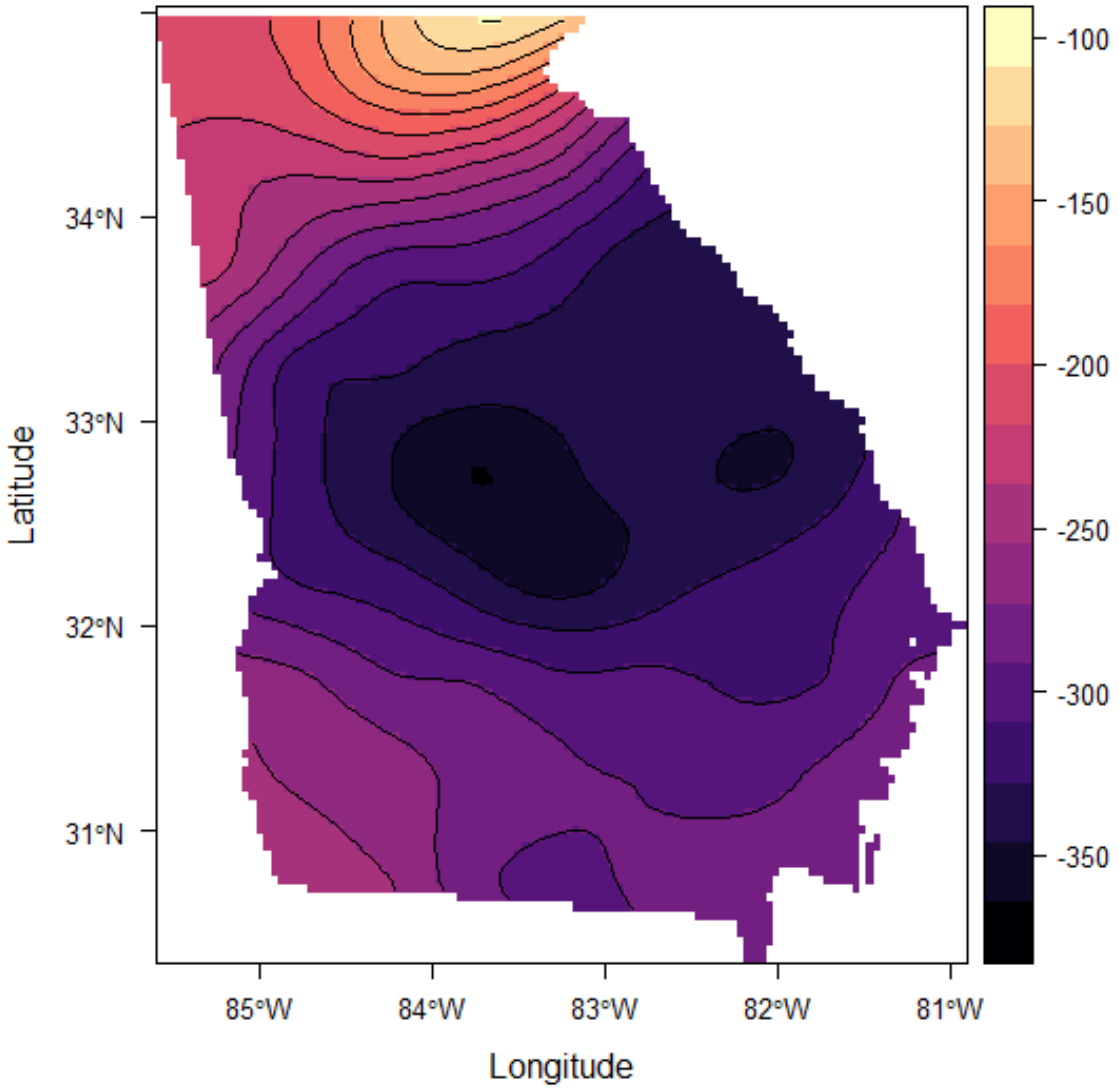

Figure 4.11: Map of water deficit (in millimeters) for Georgia, United States.

Table 4.1: Stand description for the twenty sites used in the longleaf pine study. All sites were located in South Georgia.

Stand #	County	Lat.	Lon.	Origin	Age	Est. Year
1	Dodge	32.1386	-82.9812	cut-over	16	2003
2	Dodge	32.1286	-82.983	old-field	18	2001
4	Dodge	32.1086	-83.2476	cut-over	12	2007
5	Dodge	32.0823	-83.2877	old-field	12	2007
6	Wilcox	32.0431	-83.5077	cut-over	12	2007
7	Wilcox	31.9802	-83.4756	old-field	18	2001
8	Berrien	31.346	-83.1759	old-field	20	1999
9	Berrien	31.3487	-83.1778	old-field	20	1999
10	Berrien	31.4278	-83.2774	old-field	15	2004
11	Burke	33.1279	-81.7293	cut-over	16	2003
12	Burke	33.1308	-81.7435	cut-over	19	2000
13	Stewart	32.1627	-84.9723	cut-over	15	2004
14	Charlton	30.4488	-82.0834	cut-over	25	1994
15	Brantley	31.1896	-81.8029	cut-over	19	2000
18	Long	31.6985	-81.6093	cut-over	22	1997
20	Screven	32.7196	-81.5711	old-field	19	2000
21	Screven	32.7741	-81.533	old-field	19	2000
22	Screven	32.5242	-81.5562	old-field	14	2005
23	Baker	31.1893	-84.4805	old-field	17	2002
24	Baker	31.2393	-84.4166	old-field	18	2001

Table 4.2. Summary statistics for the taper models tested.

Model	Bias	RMSE
Kozak	-1.47E-17	0.017603
Modified Kozak	0.001351	0.017114
Ormerod	0.594668	0.227632
Max and Burkhart	-1.77E-06	0.000675

Table 4.3: Minimum and maximum values for the parameters calculated using the Kozak taper model.

	Min	Max
B_0	1.195	1.408
B_1	-2.454	-1.766
B_2	0.5918	1.059